THE ANNALS *of* MATHEMATICAL STATISTICS

FOUNDED AND EDITED BY H. C. CARVER, [illegible]

EDITED BY S. S. WILKS, [illegible]

THE OFFICIAL JOURNAL OF THE INSTITUTE OF MATHEMATICAL STATISTICS

Contents

(Continued on [illegible])

Vol. 29, No. 4 — December, 1958

THE ANNALS OF MATHEMATICAL STATISTICS

Subscription Rates. Current issues are $12 per volume (four issues of one calendar year) in the U. S. and Canada, $10 per volume elsewhere. Single issues are $3.50. Back numbers are $12 per volume, $3.50 per issue, $200 for the first 25 volumes, $10 per additional volume purchased at the same time as Volume I through XXV.

Rates to members of the Institute of Mathematical Statistics are lower (see inside back cover).

Communications concerning subscriptions, back numbers, payment of dues, etc., should be addressed to A. H. Bowker, Treasurer, Department of Statistics, Stanford University, Stanford, Calif.

Communications concerning membership, changes of address, etc., should be addressed to George E. Nicholson, Jr., Secretary, Department of Statistics, University of North Carolina, Chapel Hill, N. C. Changes of address which are to become effective for a given issue of the *Annals* should be reported to the Secretary on or before the 10th of the month preceding the month of issue.

Editorial Office. Department of Statistics, Eckhart Hall, University of Chicago, Chicago 37, Illinois, W. H. Kruskal, Editor.

Preparation of manuscripts. Manuscripts should be submitted to the editorial office. Each manuscript should be typewritten, double spaced, with wide margins at sides, top, and bottom, and the original should be submitted with one additional copy, on paper that will take corrections. Dittoed or mimeographed papers are acceptable only if completely legible. Footnotes should be reduced to a minimum, and where possible replaced by remarks in the text, or a bibliography at the end of the paper; formulae in footnotes should be avoided. References should follow current *Annals* style, and should be numbered alphabetically according to authors' names.

Figures, charts, and diagrams should be professionally drawn on plain white paper or tracing cloth in black India ink twice the size they are to be printed.

Authors are asked to keep in mind the typographical difficulties of complicated mathematical formulae. The difference between capital and lower-case letters should be clearly shown; care should be taken to avoid confusion between such pairs as zero and the letter O, the numeral 1 and the letter l, numeral 1 used as superscript and prime ('), alpha and a, kappa and k, mu and u, nu and v, eta and n. The subscripts or superscripts should be clearly below or above the line. Bars over groups of letters (e.g., $\overline{\log x}$) and underlined letters (e.g., x) are difficult to print and should be avoided. Symbols are automatically italicized by the printer and should not be underlined on manuscripts. Boldface letters may be indicated by underlining with a wavy line on the manuscript; boldface subscripts and superscripts are not available. Complicated exponentials should be represented with the symbol exp. In writing square roots the fractional exponent is preferable to the radical sign. Fractions *in the body of the text* (as opposed to displayed expressions) and fractions occurring in the numerators or denominators of fractions are preferably written with the solidus, thus $(a + b)/(c + d)$ rather than $\frac{a+b}{c+d}$.

Authors will ordinarily receive only galley proofs. Fifty reprints without covers will be furnished free. Additional reprints and covers will be furnished at cost.

Entered as second-class matter at the Post Office at Baltimore, Maryland, under the act of March 3, 1879.

Composed and Printed at the
WAVERLY PRESS, Inc., Baltimore, Maryland, U. S. A.

Published quarterly by the Institute of Mathematical Statistics in March, June, September and December at Baltimore, Maryland.

ON MIXED SINGLE SAMPLE EXPERIMENTS[1]

By Leonard Cohen

College of the City of New York

1. Introduction and summary. William Kruskal [1], Howard Raiffa [2], J. L. Hodges, Jr. and E. L. Lehmann [4], have shown that in certain Neyman-Pearson type problems of testing a simple hypothesis against a simple alternative, determining the sample size by means of a chance device yields improvements over fixed sample size procedures. The purpose of this paper is not only to investigate the general problem of randomizing over fixed sample size tests of a simple hypothesis against a simple alternative, but also randomizing over other fixed sample size procedures in topics such as confidence interval estimation, the k-decision problem, etc.

In Section 2, a fixed sample size test of a simple hypothesis against a simple alternative is identified with an operating characteristic (α, β, n) where α denotes the probability of a type I error, β denotes the probability of a type II error, and n denotes the sample size. A mixed single sample test is defined as a sequence of quadruples.

$(\gamma_i, \alpha_i, \beta_i, n_i)$, where $\gamma_i \geqq 0, \sum_{i=1}^{\infty} \gamma_i = 1$, where $(\alpha_i, \beta_i, \eta_i)$ is a fixed sample size test and where γ_i is interpreted as the probability of using the fixed sample size test (α_i, β_i, n_i) for $i = 1, 2, \cdots$. A mixed single sample test is identified with an operating characteristic $(\alpha, \beta, n) = \sum_{i=1}^{\infty} \gamma_i(\alpha_i, \beta_i, n_i)$. For each non-negative integer n, the class A_n of admissible fixed sample size procedures of sample size n is defined in an obvious way. We define $A = \bigcup_{i=0}^{\infty} A_i$ and A^* as the convex hull of A. It is not necessarily true that A^* is closed. An example is given to show this. However, it is true that the lower boundary of A^* is a subset of A^* so that the lower boundary of A^* determines a minimally complete class, $\mathfrak{A}$, of mixed single sample tests. The tests in $\mathfrak{A}$ are characterized from a Bayes point of view and a technique for constructing the tests in $\mathfrak{A}$ is given.

In Section 3, the technique is applied to tests on the mean of a normal distribution with known variance. It is shown that the tests in $\mathfrak{A}$ are either

(a) fixed sample size tests, or

(b) mixtures of at most two fixed sample size tests.

It is shown that there exists a minimal subset $\mathfrak{A}_0$ of A such that all improved randomized procedures are of the form $(\alpha, \beta, n) = \gamma(0, 1, 0) + (1 - \gamma)(\alpha_0, \beta_0, n_0)$ or $(\alpha, \beta, n) = \gamma(1, 0, 0) + (1 - \gamma)(\alpha_0, \beta_0, n_0)$, where $0 < \gamma < 1$ and where $(\alpha_0, \beta_0, n_0) \varepsilon \mathfrak{A}_0$. It is then shown how to construct $\mathfrak{A}_0$. The following problems (of the Neyman-Pearson type) are solved:

(a) Given α and β, how can we find the test in $\mathfrak{A}$ with the given α and β?

Received January 2, 1958.

[1] Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, in the Faculty of Political Science, of Columbia University, 1957.

(b) Given α and n, how can we find the test in $\mathfrak{A}$ with the given α and n? Numerical examples are worked out.

In Section 4, the technique is applied to tests on the mean of a binomial distribution. Although no general results were obtained, numerical examples of interest are given.

In Section 5, the technique is applied to tests on the range of a rectangular distribution (when one end point is known). It is shown that if $\alpha > 0$, $n > 0$, and $(\alpha, \beta, n) \varepsilon A_n$, then $(\alpha, \beta, n) \not\varepsilon \mathfrak{A}$. The tests in $\mathfrak{A}$ are characterized by a simple equation which makes it easy to

(a) determine whether a given point (α, β, n) belongs to $\mathfrak{A}$, and

(b) construct any test in $\mathfrak{A}$, given two of the three coordinates.

It is shown that if $(\alpha, \beta, n) \varepsilon A_n$, then there exists a test (α, β, n') in $\mathfrak{A}$ such that $n' = (1 - \alpha)n$. Hence, the fractional saving in the expected sample size achieved by randomization is equal to α.

In Section 6, it is shown that in tests on the mean of a rectangular distribution (with known range), it never pays to randomize.

In Section 7, confidence intervals are evaluated in terms of confidence coefficient (α), expected length (L) and expected sample size (n). For the problem of obtaining a confidence interval for the mean of a normal distribution with known variance, "improved" randomized procedures exist and are of the form $(\alpha, L, n) = \gamma(0, 0, 0) + (1 - \gamma)(\alpha', L', n')$ where $0 < \gamma < 1$ and where (α', L', n') is a fixed sample size confidence interval procedure. Clearly, the randomized procedures obtained are of such a nature that the question of confidence intervals evaluated in terms of expected length and/or expected sample size is thrown open to discussion.

In Section 8, the k-decision problem is discussed. It is shown that improvements can be obtained by randomization.

In Section 9, the problem of applying mixed single sample tests of a composite hypothesis against a composite alternative is discussed.

In Section 10, mixed single sample procedures are compared to Wald's sequential probability ratio test in the problem of tests on the range of a rectangular distribution when one endpoint is known and are shown to be efficient in a certain sense.

In Section 11, the estimation problem is mentioned. It is shown that in most practical problems, fixed sample size procedures are optimal.

In Section 12, applications of mixed single sample tests are discussed.

2. Testing a simple hypothesis against a simple alternative. Let X denote a random variable with density function (or discrete probability function) $f(x, \theta)$. We wish to test the hypothesis $H_0: \theta = \theta_0$ against the alternative $H_1: \theta = \theta_1$. In the sequel, we shall restrict ourselves exclusively to fixed sample size tests, both randomized and non-randomized, and mixtures of such tests. Any test of the preceding kinds will be identified with an operating characteristic (α, β, n), where α denotes the probability of a type I error, β denotes the probability of a

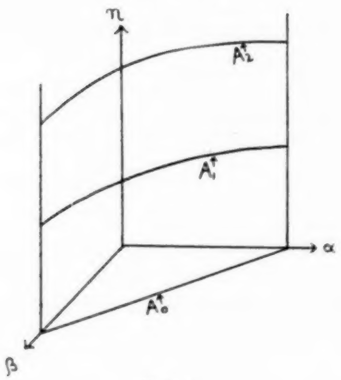


FIG. 1

type II error, and n denotes the expected number of observations. If two tests have the same operating characteristic, they will be considered equivalent.

Let $(x_1, x_2, \cdots, x_n)$ denote a sample of n independent observations on X. Let δ_n denote a real-valued, measurable function of n variables whose range is the closed interval $(0, 1)$. The expression $\delta_n(x_1, x_2, \cdots, x_n)$ is interpreted as the probability of rejecting H_0 if $(x_1, x_2, \cdots, x_n)$ is observed. Let Δ_n denote the class of functions $\{\delta_n\}$ of the preceding type.

Definition 1. For any integer $n > 0$, let $S_n = \{(\alpha, \beta, n): \alpha = E(\delta_n \mid \theta_0), \beta = E(1 - \delta_n \mid \theta_1), \delta_n \, \varepsilon \, \Delta_n\}$. S_n is the class of tests of fixed sample size n. We define $S_0 = \{(\alpha, \beta, 0): 0 \leqq \alpha \leqq 1, \alpha + \beta = 1\}$.

Definition 2. For any integer $n \geqq 0$, let $A_n = \{(\alpha, \beta, n): a)(\alpha, \beta, n) \, \varepsilon \, S_n$, and $b)$ there exists no other test (α', β', n) belonging to S_n with the property that $\alpha' \leqq \alpha$, $\beta' \leqq \beta$, at least one of these inequalities being strict.$\}$
The set A_n is the class of admissible procedure based on samples of fixed size n, and is known to be complete. See Fig. 1.

Definition 3. Let $A = \bigcup_{i=0}^{\infty} A_i$.

Definition 4. Let $A^* = \{(\alpha, \beta, n): (\alpha, \beta, n) = \sum_{i=0}^{\infty} \gamma_i(\alpha_i, \beta_i, n_i)$ where $\gamma_i \geqq 0$, $\sum_{i=0}^{\infty} \gamma_i = 1$, and $(\alpha_i, \beta_i, n_i) \, \varepsilon \, A$ for $i = 0, 1, 2, \cdots\}$.
γ_i is interpreted as the probability of selecting the fixed sample size test (α_i, β_i, n_i) for $i = 0, 1, 2, \cdots$. A^* is the convex hull of A.

Definition 5. Let $\mathfrak{A} = \{(\alpha, \beta, n): a)(\alpha, \beta, n) \, \varepsilon \, A^*$, and $b)$ there exists no other test (α', β', n') belonging to A^* with the property that $\alpha' \leqq \alpha$, $\beta' \leqq \beta$, $n' \leqq n$, at least one of these inequalities being strict.$\}$.
The set $\mathfrak{A}$ is the class of admissible mixed single sample tests.

We next wish to show that $\mathfrak{A}$ is complete, i.e., for any test (α', β', n') not in $\mathfrak{A}$, there exists a test (α, β, n) in $\mathfrak{A}$ such that $\alpha \leqq \alpha'$, $\beta \leqq \beta'$, $n \leqq n'$, at least one of these inequalities being strict. If, in general, A^* were closed, it would follow that $\mathfrak{A}$ is complete. However, A^* is not necessarily closed, as the following example will illustrate.

Example. Let $f(x, \theta) = 1$ if $\theta \leqq X \leqq \theta + 1$

$= 0$ elsewhere.

We wish to test the hypothesis $H_0:\theta = 0$ against the alternative $H_1:\theta = \theta_1$, where $0 < \theta_1 < 1$. A simple calculation shows that

$$(1) \qquad A_n = \{(\alpha, \beta, n):0 \leqq \alpha \leqq (1 - \theta_1)^n, \alpha + \beta = (1 - \theta_1)^n\}.$$

We define a sequence $\{(\alpha_k, \beta_k, n_k)\}$, where $(\alpha_k, \beta_k, n_k) = (1 - 1/k)(0, 1, 0) + (1/k)(0, (1 - \theta_1)^k, k)$. Clearly $\lim_{k\to\infty}(\alpha_k, \beta_k, n_k) = (0, 1, 1)$. However, $(0, 1, 1) \not\varepsilon A^*$. To prove this, assume $(0, 1, 1) \varepsilon A^*$. Then, since A^* is a three-dimensional convex set, $(0, 1, 1)$ can be expressed as a convex linear combination of at most four points in A, i.e., $(0, 1, 1) = \sum_{i=1}^4 \gamma_i(\alpha_i, \beta_i, n_i)$, where $\gamma_i \geqq 0, \sum_{i=1}^4 \gamma_i = 1$ and $(\alpha_i, \beta_i, n_i) \varepsilon A$ for $i = 1, 2, 3, 4$. Since $\sum_{i=1}^4 \gamma_i\beta_i = 1$, it follows that $\beta_i = 1$ if $\gamma_i > 0$. However, if $\beta_i = 1$, it follows from (1) that $n_i = 0$, contradicting the assumption $\sum_{i=1}^4 \gamma_i n_i = 1$. Q.E.D.

In order to show that $\mathfrak{A}$ is complete, we define $A_L^* = \{(\alpha, \beta, n)$: (a) (α, β, n) is a boundary point of A^*, and (b) there exists no test (α', β', n') belonging to A^* such that $\alpha' \leqq \alpha, \beta' \leqq \beta, n' \leqq n$, at least one of these inequalities being strict.$\}$.

The set A_L^* is the "lower" boundary of A^*. Clearly, $\mathfrak{A} \subset A_L^*$. We shall now prove an important theorem.

THEOREM 1. $A_L^* \subset A^*$.

PROOF. Suppose $(\alpha, \beta, n) \varepsilon A_L^*$. Then, since (α, β, n) is a boundary point of A^*, there exists a sequence of points $\{(\alpha_k, \beta_k, n_k)\}$ belonging to A^* such that

$$(\alpha, \beta, n) = \lim_{k\to\infty}(\alpha_k, \beta_k, n_k).$$

Since A^* is a three dimensional convex set, each point (α_k, β_k, n_k) of this sequence can be expressed as a convex linear combination of at most four points in A: i.e., for each k, there exist numbers $\gamma_{ik}, \alpha_{ik}, \beta_{ik}, n_{ik}$ such that $(\alpha_k, \beta_k, n_k) = \sum_{i=1}^4 \gamma_{ik}(\alpha_{ik}, \beta_{ik}, n_{ik})$, where $\gamma_{ik} \geqq 0, \sum_{i=1}^4 \gamma_{ik} = 1$ and $(\alpha_{ik}, \beta_{ik}, n_{ik}) \varepsilon A$ for $i = 1, 2, 3, 4$. Without any loss of generality, we can assume that the sequences $\{\gamma_{ik}\}$, $\{\alpha_{ik}\}$ and $\{\beta_{ik}\}$ are convergent for $i = 1, 2, 3, 4$ as k tends to infinity. Let $\gamma_i = \lim_{k\to\infty} \gamma_{ik}, \alpha_i = \lim_{k\to\infty} \alpha_{ik}, \beta_i = \lim_{k\to\infty} \beta_{ik}$ for $i = 1, 2, 3, 4$. Clearly, $\gamma_i \geqq 0, \sum_{i=1}^4 \gamma_i = 1$ for $i = 1, 2, 3, 4$. Before proceeding with the proof of Theorem 1, we prove a useful lemma.

LEMMA 1. *If $\gamma_i > 0$, there exists a number N_i such that $n_{ik} \leqq N_i$ for all k.*

PROOF. Since $\lim_{k\to\infty} n_k = n$, there exists a positive integer K such that if $k > K$, $n_k < n + 1$. Furthermore, since $\lim_{k\to\infty} \gamma_{ik} = \gamma_i > 0$, there exists a number K_i such that if $k > K_i, \gamma_{ik} > \frac{1}{2}\gamma_i$. Let $M_i = \max(K, K_i)$. Then, if $k > M_i, \frac{1}{2}\gamma_i n_{ik} \leqq \sum_{i=1}^4 \gamma_{ik} n_{ik} = n_k < n + 1$. Thus, if $k > M_i, n_{ik} < 2(n + 1)/\gamma_i$. Then, $N_i = \max[n_{i1}, n_{i2}, \cdots n_{iM_i}, 2(n + 1)/\gamma_i]$ is the required number, proving the lemma.

We now proceed with the proof of Theorem 1. Consider four cases.

Case 1. $\gamma_i > 0, i = 1, 2, 3, 4$.

Let $N = \max(N_1, N_2, N_3, N_4)$ where N_i is defined in Lemma 1. Then, since

$0 \leqq n_{ik} \leqq N$, for $i = 1, 2, 3, 4$ and all k, the sequences $\{n_{ik}\}$ are bounded. Hence, for each i, there exists a convergent subsequence which we denote by $\{\bar{n}_{ik}\}$. Let $\{\bar{\alpha}_{ik}\}$, $\{\bar{\beta}_{ik}\}$ and $\{\bar{\gamma}_{ik}\}$ denote respectively the subsequences of $\{\alpha_{ik}\}$, $\{\beta_{ik}\}$ and $\{\gamma_{ik}\}$ corresponding to the convergent subsequence $\{\bar{n}_{ik}\}$ of $\{n_{ik}\}$. Let $\lim_{k\to\infty} \bar{n}_{ik} = n_i$. Clearly,

$$(\alpha, \beta, n) = \lim_{k\to\infty} (\alpha_k, \beta_k, n_k) = \lim_{k\to\infty} \sum_{i=1}^{4} \gamma_{ik}(\alpha_{ik}, \beta_{ik}, n_{ik}) = \sum_{i=1}^{4} \lim_{k\to\infty} \gamma_{ik}(\alpha_{ik}, \beta_{ik}, n_{ik})$$

$$= \sum_{i=1}^{4} \lim_{k\to\infty} \bar{\gamma}_{ik}(\bar{\alpha}_{ik}, \bar{\beta}_{ik}, \bar{n}_{ik}) = \sum_{i=1}^{4} \gamma_i(\alpha_i, \beta_i, n_i).$$

Since $(\bar{\alpha}_{ik}, \bar{\beta}_{ik}, \bar{n}_{ik}) \varepsilon A$ for all i and k, and since A is closed,

$$\lim_{k\to\infty}(\bar{\alpha}_{ik}, \bar{\beta}_{ik}, \bar{n}_{ik}) = (\alpha_i, \beta_i, n_i) \epsilon A \qquad \textit{for } i = 1, 2, 3, 4.$$

Furthermore, since $\gamma_i > 0$, $i = 1, 2, 3, 4$, and $\sum_{i=1}^{4} \gamma_i = 1, \sum_{i=1}^{4} \gamma_i(\alpha_i, \beta_i, n_i) \varepsilon A^*$. Hence $(\alpha, \beta, n) \varepsilon A^*$.

The more difficult case to prove is Case 2.

Case 2. Exactly one of the γ_i's is 0.

To fix ideas, suppose $\gamma_1 = 0$, $\gamma_2 > 0$, $\gamma_3 > 0$, $\gamma_4 > 0$. Let $N = \max(N_2, N_3, N_4)$. In a manner analogous to that used in Case 1, we define sequences $\{\bar{\alpha}_{ik}\}$, $\{\bar{\beta}_{ik}\}$, $\{\bar{n}_{ik}\}$ and $\{\bar{\gamma}_{ik}\}$ for $i = 2, 3, 4$. We define ne w sequences

$$\alpha_k' = \bar{\gamma}_{1k}(0) + \sum_{i=2}^{4} \bar{\gamma}_{ik} \bar{\alpha}_{ik},$$

$$\beta_k' = \bar{\gamma}_{1k}(1) + \sum_{i=2}^{4} \bar{\gamma}_{ik} \bar{\beta}_{ik},$$

$$n_k' = \bar{\gamma}_{1k}(0) + \sum_{i=2}^{4} \bar{\gamma}_{ik} \bar{n}_{ik},$$

where $\bar{\gamma}_{1k} = 1 - \sum_{i=2}^{4} \bar{\gamma}_{ik}$. It is easily seen that

$$\lim_{k\to\infty} \alpha_k' = \alpha,$$

$$\lim_{k\to\infty} \beta_k' = \beta,$$

$$\lim_{k\to\infty} n_k' \leqq n,$$

Since $(\alpha_k', \beta_k', n_k') \varepsilon A^*$ for each k, and since $(\alpha, \beta, n) \varepsilon A_L^*$, it follows that the inequality $\lim_{k\to\infty} n_k' < n$ cannot hold. Hence,

$$(\alpha, \beta, n) = \lim_{k\to\infty} (\alpha_k', \beta_k', n_k') = \lim_{k\to\infty} \sum_{i=2}^{4} \bar{\gamma}_{ik},(\bar{\alpha}_{ik}, \bar{\beta}_{ik}, \bar{n}_{ik})$$

$$\cdot \sum_{i=2}^{4} \lim \bar{\gamma}_{ik}(\bar{\alpha}_{ik}, \bar{\beta}_{ik}, \bar{n}_{ik}) = \sum_{i=2}^{4} \gamma_i(\alpha_i, \beta_i, n_i).$$

Using the argument in Case 1, we find that $(\alpha, \beta, n) \varepsilon A^*$.

Case 3. Two of the γ_i's are 0. The proof of Case 3 is analogous to the proof of Case 2.

Case 4. Three of the γ_i's are 0. The proof of Case 4 is analogous to the proof of Case 2.

COROLLARY 1. $\mathfrak{A} = A_L^*$.

COROLLARY 2. $\mathfrak{A}$ *is complete.*

PROOF. Let (α', β', n') be a test which does not belong to $\mathfrak{A}$.

Let

$$A_{(\alpha'\beta')} = \{(\alpha, \beta, n): a)\ \alpha = \alpha', \beta = \beta' \text{ and } (\alpha, \beta, n) \varepsilon A^*\}.$$

$A_{(\alpha',\beta')}$ is non-empty since $(\alpha', \beta', n') \varepsilon A_{(\alpha',\beta')}$. Let

$$N = N_{(\alpha',\beta')} = \inf_{\{n':(\alpha',\beta',n')\varepsilon A_{(\alpha',\beta')}\}} n'$$

Then $(\alpha', \beta', N) \varepsilon \mathfrak{A} = A_L^*$ where $N < n'$.

Note: It is possible to show that $\mathfrak{A}$ is complete using a different approach. If we define $S = \bigcup_{i=0}^* S_i$ and S^* as the convex hull of S, it can be shown that S^* is closed. This implies that $\mathfrak{A}$ is complete. However, to prove that S^* is closed requires a technique similar to that used in proving Theorem 1.

THEOREM 2. *If $f(x, \theta_0) = 0$ if and only if $f(x, \theta_1) = 0$, then a necessary and sufficient condition for (α, β, n) to belong to $\mathfrak{A}$ is that for some non-negative a and b and positive c, we have*

$$a\alpha + b\beta + cn = \min_{(\alpha',\beta',n')\varepsilon A^*} \{a\alpha' + b\beta' + cn'\}.$$

PROOF. To prove the sufficiency of the condition, we consider 4 cases.

Case 1. $a = 0, b = 0, c > 0$. Then $a\alpha' + b\beta' + cn' = cn'$ is minimized *only* by tests $(\alpha, \beta, 0)$ belonging to A_0. However, $A_0 \subset \mathfrak{A}$, proving the sufficiency of the condition if Case 1 holds.

Case 2. $a = 0, b > 0, c > 0$. Then, $a\alpha' + b\beta' + cn' = b\beta' + cn'$ is minimized only by the test $(1, 0, 0)$ which belongs to A_0.

Case 3. $a > 0, b = 0, c > 0$. (Similar to Case 2.)

Case 4. $a > 0, b > 0, c > 0$. Then, it is well known, and can be easily proved that any test (α, β, n) such that $a\alpha + b\beta + cn = \min_{(\alpha',\beta',n')\varepsilon A^*}(a\alpha' + b\beta' + cn')$ belongs to $\mathfrak{A}$.

To prove the necessity of the condition, we assume $(\alpha, \beta, n) \varepsilon \mathfrak{A}$.

(i) If $n = 0$, choose $a = 0, b = 0, c = 1$.

(ii) If $n > 0$, then it is well known in the theory of convex sets that there exist non-negative numbers a, b and c such that

$$a\alpha + b\beta + cn = \min_{(\alpha',\beta',n')\varepsilon A^*} (a\alpha' + b\beta' + cn').$$

It remains to show that $c > 0$. Assume $c = 0$. Then

$$a\alpha + b\beta = \min_{(\alpha',\beta',n')\varepsilon A^*} (a\alpha' + b\beta') = 0.$$

Since $(\alpha, \beta, n) \varepsilon \mathfrak{A}$, then there exist numbers $\gamma_i, \alpha_i, \beta_i, n_i$ such that $(\alpha, \beta, n) = \sum_{i=1}^{4} \gamma_i(\alpha_i, \beta_i, n_i)$, where $\gamma_i \geqq 0, \sum_{i=1}^{4} \gamma_i = 1$, $(\alpha_i, \beta_i, n_i) \varepsilon A$ for $i = 1, 2, 3, 4$.

Thus,

$$a\alpha + b\beta = a \sum_{i=1}^{4} \gamma_i \alpha_i + b \sum_{i=1}^{4} \gamma_i \beta_i = 0.$$

Since both a and b cannot equal 0, either $\alpha = 0$ or $\beta = 0$. Assume $\alpha = 0$. Then, if $\gamma_i > 0$, $\alpha_i = 0$. Using the fact that $f(x, \theta_0) = 0$ if and only if $f(x, \theta_1) = 0$, it follows that if $\alpha_i = 0$, $\beta_i = 1$. Hence, $(\alpha, \beta, n) = (0, 1, n)$. But, $(0, 1, n) \not\varepsilon \mathfrak{A}$ since $(0, 1, 0)$ is preferred. Thus we are led to a contradiction of the fact that $(\alpha, \beta, n) \varepsilon \mathfrak{A}$. If we assume $\beta = 0$, we are led to a similar contradiction. Therefore, the assumption $c = 0$ is false. Theorem 2 is thus proved.

Theorem 2 states, in effect, that the problem of generating $\mathfrak{A}$ reduces to constructing tests (α, β, n) which minimize the expression $a\alpha + b\beta + cn$ for all choices of non-negative a and b and positive c. The cases where either a or b is 0 were discussed and disposed of in proving Theorem 2. The main problem, then, is to construct the tests (α, β, n) which minimize the expression $a\alpha + b\beta + cn$. We proceed as follows: without any loss of generality we may assume that $a + b = 1$ and write $a = \pi$ and $b = 1 - \pi$, where $0 < \pi < 1$. Then, we wish to find the tests (α, β, n) in $\mathfrak{A}$ such that

$$\pi\alpha + (1 - \pi)\beta + cn = \min_{(\alpha', \beta', n') \varepsilon A^*} [\pi\alpha' + (1 - \pi)\beta' + cn'].$$

Clearly,

$$\begin{aligned}
&\min_{(\alpha', \beta', n') \varepsilon A^*} [\pi\alpha' + (1 - \pi)\beta' + cn'] \\
&= \min_{\substack{[\gamma_i, \alpha_i, \beta_i, n_i : \gamma_i \geqq 0, \sum_{i=0}^{\infty} \gamma_i = 1, (\alpha_i, \beta_i, n_i) \varepsilon A] \\ i = 0, 1, 2, \cdots}} \cdot \left\{\pi \sum_{i=0}^{\infty} \gamma_i \alpha_i + (1 - \pi) \sum_{i=0}^{\infty} \gamma_i \beta_i + c \sum_{i=0}^{\infty} \gamma_i n_i\right\} \\
&= \min_{\substack{[\gamma_i, \alpha_i, \beta_i, n_i : \gamma_i \geqq 0, \sum_{i=0}^{\infty} \gamma_i = 1, (\alpha_i, \beta_i, n_i) \varepsilon A] \\ i = 0, 1, 2, \cdots}} \cdot \left\{\sum_{i=0}^{\infty} \gamma_i[\pi\alpha_i + (1 - \pi)\beta_i] + c \sum_{i=0}^{\infty} \gamma_i n_i\right\} \\
&= \min_{N \geqq 0} \left(cN + \left\{ + \min_{(\gamma_i, n_i : \sum_{i=0}^{\infty} \gamma_i n_i = N, \gamma_i \geqq 0, \sum \gamma_i = 1, i = 0, 1, 2, \cdots)} \cdot \sum_{i=0}^{\infty} \gamma_i \min_{(\alpha_i, \beta_i, n_i) \varepsilon A_{n_i}} [\pi\alpha_i + (1 - \pi)\beta_i]\right\}\right).
\end{aligned}$$

It should be noted that the operation "min"$_{N \geqq 0}$ is not restricted to integral values of N.

From the above, it is clear that the desired minimization can be accomplished in 3 steps, which we shall now describe in detail.

Step 1. We can, for each n_i, find the tests (α_i, β_i, n_i) belonging to A_{n_i} which minimize the expression $\pi\alpha_i + (1 - \pi)\beta_i$. For each n_i, let

$$R_\pi(n_i) = \min_{\{(\alpha_i,\beta_i):(\alpha_i,\beta_i,n_i)\varepsilon A_{n_i}\}} \{\pi\alpha_i + (1 - \pi)\beta_i\}.$$

$R_\pi(n_i)$ may be interpreted as the Bayes risk for fixed sample-size procedures of sample size n_i where π is the a priori probability that θ_0 is the true parameter and $1 - \pi$ is the a priori probability that θ_1 is the true parameter.

In particular,

$$R_\pi(0) = \min_{\{\alpha,\beta:0\leq\alpha\leq 1,\alpha+\beta=1\}} \{\pi\alpha + (1 - \pi)\beta\} = \min(\pi, 1 - \pi).$$

If $0 < \pi < \frac{1}{2}$, $R_\pi(0) = \pi$. The only test $(\alpha, \beta, 0)$ belonging to A_0 satisfying the equation $\pi\alpha + (1 - \pi)\beta = \pi$ is the test $(1, 0, 0)$. Similarly, if $\frac{1}{2} < \pi < 1$, $R_\pi(0) = 1 - \pi$. The only test belonging to A_0 satisfying the equation $\pi\alpha + (1 - \pi)\beta = 1 - \pi$ is the test $(0, 1, 0)$. If $\pi = \frac{1}{2}$, $R_\pi(0) = \frac{1}{2}$. Then, any test belonging to A_0 satisfies the equation $\frac{1}{2}\alpha + \frac{1}{2}\beta = \frac{1}{2}$, since $\alpha + \beta = 1$.

We note that

$$\min_{(\alpha',\beta',n')\varepsilon A^*} [\pi\alpha' + (1 - \pi)\beta' + cn']$$

$$= \min_{N\geq 0}\left\{cN + \min_{[\gamma_i,n_i:\gamma_i\geq 0,\sum_{i=0}^\infty\gamma_i=1,\sum_{i=0}^\infty\gamma_i n_i=N]} \sum_{i=0}^\infty \gamma_i R_\pi(n_i)\right\}.$$

Step 2. Subject to the conditions

$$\gamma_i \geqq 0,\ i = 0, 1, 2, \cdots, \textstyle\sum_{i=0}^\infty \gamma_i = 1, \sum_{i=0}^\infty \gamma_i n_i = N,$$

we can, for each non-negative value of N choose the γ_i's so that $\sum_{i=0}^\infty \gamma_i R_\pi(n_i)$ is minimized. To this end, let

$$R_\pi = \bigcup_{k=0}^\infty (k, R_\pi(k)).$$

Let R_π^* denote the convex hull of R_π and let $\mathcal{R}_\pi$ denote the lower boundary of R_π^*, i.e., $\mathcal{R}_\pi = \{(k, r):$(a) $(k, r)\ \varepsilon\ R_\pi^*$ and (b) there exists no point (k', r') belonging to R_π^* such that $k' \leqq k$, $r' \leqq r$, at least one of these inequalities being strict.$\}$.

Then, to accomplish Step 2 of the minimization, given $N \geqq 0$, we merely select the point (N, r) belonging to $\mathcal{R}_\pi$. Since (N, r) is a boundary point of a two dimensional convex set, (N, r) *can always be expressed as a convex linear combination of at most two points in* R. We define

$$r_\pi(N) = \min_{[\gamma_i,n_i:\gamma_i\geq 0,\sum_{i=0}^\infty\gamma_i=1,\sum_{i=0}^\infty\gamma_i n_i=N,i=0,1,2,\cdots]} \left\{\sum_{i=0}^\infty \gamma_i R_\pi(n_i)\right\}.$$

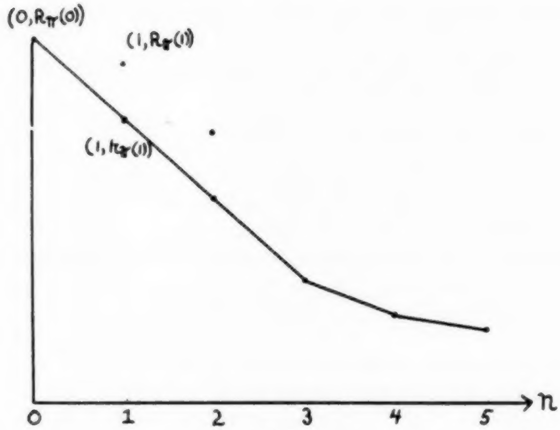


FIG. 2

See Fig. 2. We note that

$$\min_{(\alpha',\beta',n') \in A^*} [\pi\alpha' + (1 - \pi)\beta' + cn'] = \min_{N \geqq 0} [r_\pi(N) + cN].$$

Step 3. We now wish to choose $N \geqq 0$ to minimize the expression $r_\pi(N) + cN$. Since $r_\pi(N)$ is a strictly decreasing, convex and piecewise linear function of N, there exists at least one value of N and at most a finite interval of values of N which minimize $r_\pi(N) + cN$.

It should be noted that if we are given a specific value of N, then there exists a number $c > 0$ such that $r_\pi(N) + cN = \min_k [r_\pi(k) + ck]$. Therefore, for an arbitrary but fixed value of $N > 0$ any procedure obtained in Step 2 will be an admissible mixed single sample test so that Step 3 is inessential in constructing $\mathcal{A}$.

We shall apply the technique in several problems in the following sections.

3. Testing the mean of a normal distribution when the variance is known. Let

$$f(x, \theta) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2}\left(\frac{x-\theta}{\sigma}\right)^2\right\},$$

where $\sigma > 0$ is known. We wish to test the hypothesis $H_0:\theta = \theta_0$ against the alternative $H_1:\theta = \theta_1$, $\theta_0 < \theta_1$. It can be shown that for any integer $n \geqq 0$,

$$A_n = \{(\alpha, \beta, n):\alpha = 1 - \Phi(t), \beta = \Phi(t - \sqrt{n}\delta) \text{ for } -\infty \leqq t \leqq \infty\},$$

where

$$\Phi(t) = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}\, dx$$

and

$$\delta = \frac{\theta_1 - \theta_0}{\sigma}.$$

Step 1. We have already seen that $R_\pi(0) = \min(\pi, 1 - \pi)$. For any integer $n > 0$,

$$
\begin{aligned}
R_\pi(n) &= \min_{\{(\alpha,\beta):(\alpha,\beta,n)\,\varepsilon\, A_n\}} [\pi\alpha + (1-\pi)\beta] \\
&= \min_t \{\pi[1 - \Phi(t)] + (1-\pi)\Phi(t - \sqrt{n}\delta)\} \qquad (2)\\
&= \pi\left[1 - \Phi\left(\frac{\xi}{\sqrt{n}\delta} + \frac{\sqrt{n}\delta}{2}\right)\right] + (1-\pi)\,\Phi\left(\frac{\xi}{\sqrt{n}\delta} - \frac{\sqrt{n}\delta}{2}\right)\right],
\end{aligned}
$$

where $\xi = \log \pi/(1 - \pi)$. Furthermore, the test (α, β, n) such that

$$\alpha = 1 - \Phi\left(\frac{\xi}{\sqrt{n}\delta} + \frac{\sqrt{n}\delta}{2}\right) \quad \text{and} \quad \beta = \Phi\left(\frac{\xi}{\sqrt{n}\delta} - \frac{\sqrt{n}\delta}{2}\right)$$

is unique. It should also be noted that for any π such that $0 < \pi < 1$, $R_\pi(n)$ is a strictly decreasing function of n. See Figure 3.

Step 2. To accomplish Step 2 of the minimization, we consider $R_\pi(n)$ formally as a function of a continuous variable n. We shall first show that there exists a number $n_i = n_i(\pi)$ such that $R_\pi(n)$ is concave on the interval $(0, n_i)$ and convex on the interval (n_i, ∞). To show the existence of n_i, we use the identities

(a) $\varphi(x - y) = e^{2xy}\varphi(x + y)$,

(b) $\varphi'(x) = -x\varphi(x)$, where $\varphi(x) = \Phi'(x)$.

A routine calculation shows that

(c) $R'_\pi(n) = \dfrac{-\pi}{2\sqrt{n}\delta}\,\varphi\left(\dfrac{\xi}{\sqrt{n}\delta} + \dfrac{\sqrt{n}\delta}{2}\right)$

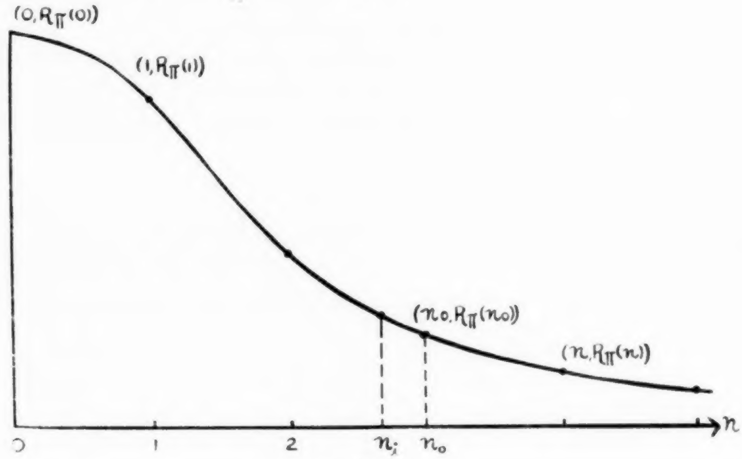


FIG. 3

and

$$\text{(d)}\quad R''_\pi(n) = (n^2\delta^4 + 4n\delta^2 - 4\xi^2)\frac{\pi\sqrt{n}\delta}{16n^3}\varphi\left(\frac{\xi}{\sqrt{n}\delta} + \frac{\sqrt{n}\delta}{2}\right).$$

Setting $R''_\pi(n)$ equal to 0, we find that

$$n_i = \frac{-2 + 2\sqrt{1+\xi^2}}{\delta^2}. \tag{3}$$

Therefore, (3) gives a unique inflection point of the function $R_\pi(n)$. See Fig. 3.

Since $R_\pi(n)$, defined in (2), is defined only for integral values of n, and since n_i in general is not an integer, we assert that there exists an integer $n_0 = n_0(\pi)$ such that $R_\pi(n)$ is concave on the interval $(0, n_0)$ and convex on the interval (n_0, ∞). See Fig. 3. It then follows that

$$r_\pi(N) = \begin{cases} \left(1 - \dfrac{N}{n_0}\right)R_\pi(0) + \dfrac{N}{n_0}R_\pi(n_0) & \text{if } N \leqq n_0 \\ ([N] + 1 - N)R_\pi([N]) + (N - [N])R_\pi([N]+1) & \text{if } N > n_0 \end{cases}$$

Thus Step 2 of the minimization is achieved.

It now becomes clear that improved randomized procedures (α, β, n) exist and are of the form $(\alpha, \beta, n) = \gamma(0, 1, 0) + (1 - \gamma)(\alpha_0, \beta_0, n_0)$ or $(\alpha, \beta, n) = \gamma(1, 0, 0) + (1 - \gamma)(\alpha_0, \beta_0, n_0)$ where $0 < \gamma < 1$ and where

$$n_0 = n_0(\pi), \qquad \alpha_0 = \alpha_0(\pi) = 1 - \Phi\left(\frac{\xi}{\sqrt{n_0}\delta} + \frac{\sqrt{n_0}\delta}{2}\right),$$

$$\beta_0 = \beta_0(\pi) = \Phi\left(\frac{\xi}{\sqrt{n_0}\delta} - \frac{\sqrt{n_0}\delta}{2}\right)$$

for some π such that $0 < \pi < 1$.

It also becomes clear that a test $(\alpha, \beta, n) \in A_n$ if and only if $n \geqq n_0(\pi)$, where π is defined by the equation

$$\beta = \Phi\left(\frac{\log \dfrac{\pi}{1-\pi}}{\sqrt{n}\delta} - \frac{\sqrt{n}\delta}{2}\right).$$

This gives a complete answer to the general question of whether or not a fixed sample size procedure can be improved upon by means of randomization.

3.1. We now consider the following problem: Given α and β, how can we find the test in $\mathfrak{A}$ achieving the given α and β? To this end, consider two cases.

Case 1. $\alpha < \beta$. Let $\mathfrak{A}_0 = \{(\alpha_0, \beta_0, n_0): n_0 = n_0(\pi), \beta_0 = \beta_0(\pi), \alpha_0 = \alpha_0(\pi) \text{ for } \frac{1}{2} < \pi < 1\}$.

Let (α, β, n) denote the test in $\mathfrak{A}$ with the given α and n.

From the discussion of Step 2, it is evident that (α, β, n) is an improved randomized procedure if and only if $(\alpha, \beta, n) = \gamma(0, 1, 0) + (1 - \gamma)(\alpha_0, \beta_0, n_0)$, where $0 < \gamma < 1$ and where $(\alpha_0, \beta_0, n_0) \in \mathfrak{A}$. In this case, $\alpha = (1 - \gamma)\alpha_0$, $\beta =$

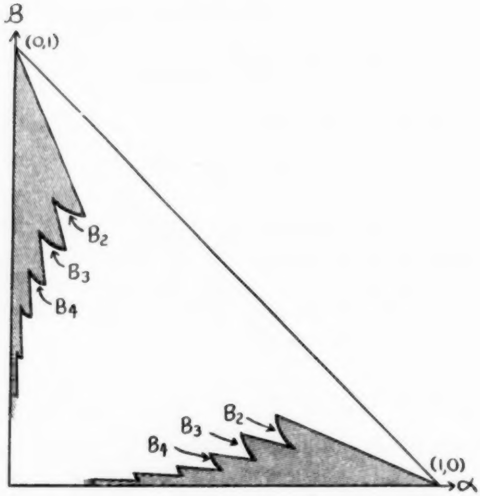


FIG. 4. Shaded region corresponds to the set of (α, β) for which the admissible test (α, β, n) is a randomized procedure; $B_i = \{(\alpha_0, \beta_0): n_0 = i\}$, $\bigcup_{i=2}^{\infty} B_i = P(\mathfrak{A}_0 \mid \alpha, \beta)$.

$\gamma + (1 - \gamma)\beta_0$, $n = (1 - \gamma)n_0$. These equations imply that $\alpha/(1 - \beta) = \alpha_0/(1 - \beta_0)$ and $1 - \gamma = \alpha/\alpha_0$. The equation $\alpha/(1 - \beta) = \alpha_0/(1 - \beta_0)$ when interpreted geometrically means that the points $(0, 1)$, (α, β) and (α_0, β_0) are colinear. The equation $1 - \gamma = \alpha/\alpha_0$ when interpreted geometrically means that (α, β) is between $(0, 1)$ and (α_0, β_0).

If (α, β, n) is not an improved randomized procedure, then

$$(\alpha, \beta, n) = \gamma(\alpha_1, \beta_1, [n]) + (1 - \gamma)(\alpha_2, \beta_2, [n] + 1)$$

where $0 \leqq \gamma \leqq 1$ and where $(\alpha_1, \beta_1, [n])$ and $(\alpha_2, \beta_2, [n] + 1) \varepsilon A$ and is of little interest.

We summarize the preceding as follows: Let $P(\mathfrak{A}_0 \mid \alpha, \beta)$ denote the projection of $\mathfrak{A}_0$ on the (α, β) plane. See Fig. 4. It was convenient to let $\delta = 1$. If (α, β) lies on a line segment joining $(0, 1)$ to one of the points (α_0, β_0) in $P(\mathfrak{A}_0 \mid \alpha, \beta)$, then the test $(\alpha, \beta, n) = (1 - \alpha/\alpha_0)(0, 1, 0) + \alpha/\alpha_0(\alpha_0, \beta_0, n_0)$ is the test in $\mathfrak{A}$ with the given α and β. Otherwise, (α, β, n) is achieved by randomizing over two fixed sample size procedures, one in $A_{[n]}$ and the other in $A_{[n]+1}$.

Case 2. $\alpha > \beta$. Similar to Case 1.

Table (1) shows the improvement in the expected sample size N which can be achieved for selected tests (α, β, n) belonging to $A_n - \mathfrak{A}$. In this case, we let $\delta = .1$.

3.2. Consider next the following problem: Given α and n, how can we construct the test in $\mathfrak{A}$ having the given α and n? We solve this problem geometri-

TABLE 1

α	β	Sample size, n, of admissible single sample tests achieving the given α and β	Expected sample size, N, of admissible mixed single sample test achieving the given α and β	Percent saving $\frac{n-N}{n} \times 100$
.005	.862	221	119	46
.005	.732	383	287	25
.01	.732	147	84	43
.01	.463	585	574	2
.05	.687	134	123	8
.05	.868	28	20	28

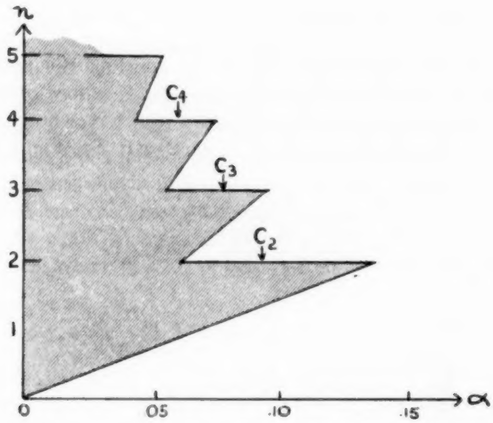


FIG. 5. Shaded region corresponds to the set of (α, n) for which the admissible test (α, β, n) is a randomized procedure; $C_i = \{(\alpha_0, n_0) : n_0 = i\}$, $\bigcup_{i=2}^{\infty} C_i = P(\mathfrak{A}_0 \mid \alpha, n)$.

cally. Let $P(\mathfrak{A}_0 \mid \alpha, n)$ denote the projection of the set $\mathfrak{A}_0$ on the (α, n) plane. See Fig. 5. Then, draw a line of slope n/α through the origin. Determine the point of intersection (α_0, n_0) of this line and $P(\mathfrak{A}_0 \mid \alpha, n)$. Clearly, $n/\alpha = n_0/\alpha_0$. If $\alpha_0 > \alpha$, the test in $\mathfrak{A}$ having the given α and n is the mixture

$$\frac{\alpha}{\alpha_0}(\alpha_0, \beta_0, n_0) + \left(1 - \frac{\alpha}{\alpha_0}\right)(0, 1, 0).$$

If $\alpha_0 \leq \alpha$, the test in $\mathfrak{A}$ having the given α and n is a mixture of two tests, one in $A_{[n]}$ and the other in $A_{[n]+1}$ and hence is of little interest.

4. Tests on the mean of a binomial distribution. Let

$$\begin{aligned} f(x, \theta) &= \theta^x(1-\theta)^{1-x} && \text{if } x = 0, 1, \\ &= 0 && \text{elsewhere, } 0 < \theta < 1. \end{aligned}$$

We wish to test the hypothesis $H_0: \theta = \theta_0$ against the alternative $H_1: \theta = \theta_1$, $\theta_1 > \theta_0$. It is known that

$$
\begin{aligned}
A_n = \Bigg\{ &(\alpha, \beta, n): \alpha = \sum_{i=0}^{n+1} \gamma_i \alpha_i, \beta = \sum_{i=0}^{n+1} \gamma_i \beta_i, \\
&\gamma_0 = \gamma_1 = \gamma_{i-1} = \gamma_{i+2} = \cdots \gamma_{n+1} = 0, \gamma_i \geqq 0, \\
&\gamma_{i+1} \geqq 0, \sum_{i=0}^{n+1} \gamma_i = 1, \\
&\alpha_i = \sum_{i=1}^{n} \binom{n}{r} \theta_0^r (1 - \theta_0)^{n-r}, \qquad \beta_i = \sum_{r=0}^{i-1} \binom{n}{r} \theta_1^r (1 - \theta_1)^{n-r}, \\
&i = 0, 1, 2, \cdots n + 1 \Bigg\}.
\end{aligned}
$$

Howard Raiffa [2] has pointed out that if we consider the projections of A_1 and A_2 on the (α, β) plane, there exists a test in A_2 whose operating characteristic is $(\theta_0, 1 - \theta_1, 2)$. However, there exists a test in A_1 whose operating characteristic is $(\theta_0, 1 - \theta_1, 1)$. Hence $(\theta_0, 1 - \theta_1, 2) \not\in \mathcal{A}$. See Fig. 6. Furthermore, if π is such that

$$
\frac{\pi}{1 - \pi} = \frac{\theta_0(1 - \theta_0)}{\theta_1(1 - \theta_1)},
$$

then $R_\pi(1) = R_\pi(2)$.

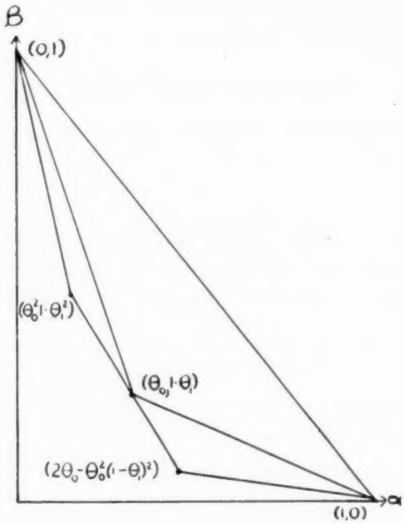


Fig. 6

TABLE 2

α	n	Probability of a type II error of admissible single sample test having the given α and n	Probability of a type II error of randomized test having the given α and n	Percent decrease
.512	30	.029	.018	38
.361	20	.112	.090	19
.098	20	.320	.286	11
.350	40	.030	.024	20

Unlike the normal distribution, there does not exist an integer $n_0(\pi)$ such that $R_\pi(n)$ is concave on the interval $(0, n_0(\pi))$ and convex on the interval $(n_0(\pi), \infty)$. Rather, it was found by numerical calculation that $R_\pi(n)$ has many inflection points. Thus, we do not generalize any further and present the following examples.

Example I. Let $\theta_0 = .04$, $\theta_1 = .15$. Table 2 shows the percent decrease in the probability of a type II error that randomization achieves over fixed sample size procedures for the given α and n. Since $R_\pi(n)$ was calculated for values of n where $n = 5k$ where k is a non-negative integer, it cannot be said with certainty that the improvements shown in Table 2 are optimal. However, the optimal improvements are at least as great as the ones recorded.

Example II. We again wish to test the hypothesis $H_0:\theta = \theta_0$ against the alternative $H_1:\theta = \theta_1$ where $\theta_0 < \theta_1$. Then, it is well known that any test $(\alpha, \beta, 1)$ such that

$$(\alpha, \beta, 1) = \gamma(0, 1, 1) + (1 - \gamma)(\theta_0, 1 - \theta_1, 1),$$

where $0 \leqq \gamma \leqq 1$ belongs to A_1. We shall now show that if we are given a test $(\alpha, \beta, 1)$ of the above type such that $\gamma \leqq 1 - \theta_1/2$ then there exists a mixed single sample test $(\alpha^*, \beta, 1)$ such that

$$\frac{\alpha - \alpha^*}{\alpha} = \frac{\gamma(1 - \alpha - \beta)}{(1 - \gamma)(1 + \beta - 2\gamma)}. \tag{4}$$

The expression $(\alpha - \alpha^*)/\alpha$ is interpreted as the fractional saving in α achieved by randomization.

To prove this, consider the test

$$(\alpha^*, \beta', n') = \frac{\gamma}{2 - \theta_1}(0, 1, 0) + \frac{\gamma}{2 - \theta_1}(\theta_0^2, 1 - \theta_1^2, 2) + \left(1 - \frac{2\gamma}{2 - \theta_1}\right)(\theta_0, 1 - \theta_1, 1).$$

Since $\gamma \leqq 1 - \theta_1/2$, the above test is a bonafide mixture. It is easily verified that $\beta' = \beta$, $n' = 1$ and that $(\alpha - \alpha^*)/\alpha$ has the value given in (4).

To illustrate the fractional saving in α which can be achieved, consider the

test $H_0:\theta = .10$ against $H_1:\theta = .95$. Then, there exists a test $(\alpha, \beta, 1)$ in A_1 where $(\alpha, \beta, 1) = .5(0, 1, 1) + .5(.10, .05, 1) = (.05, .525, 1)$. Consider the test

$$(\alpha^*, \beta, 1) = \frac{.5}{2 - .95}(0, 1, 0) + \frac{.5}{2 - .95}(.01, .0975, 2) + \left(1 - \frac{1}{2 - .95}\right)(.10, .05, 1) = \left(\frac{1}{105}, .525, 1\right).$$

Then

$$\frac{\alpha - \alpha^*}{\alpha} = \frac{17}{21}.$$

5. Tests on the range of a rectangular distribution when one endpoint is known. Let

$$f(x, \theta) = \frac{1}{\theta} \quad \text{if } 0 \leqq x \leqq \theta,$$
$$= 0 \quad \text{elsewhere.}$$

We wish to test the hypothesis $H_0:\theta = \theta_0$ against the alternative $H_1:\theta = \theta_1$, $\theta_1 > \theta_0$. It can be shown that

$$\text{(5)} \quad A_n = \left\{(\alpha, \beta, n):\alpha = \frac{\theta_0^n - t^n}{\theta_1^n}, \beta = \frac{t^n}{\theta_1^n}, 0 \leqq t \leqq \theta_0\right\} = \left\{(\alpha, \beta, n):0 \leqq \alpha \leqq 1, \beta = \left(\frac{\theta_0}{\theta_1}\right)^n (1 - \alpha)\right\}.$$

It should be noted that Theorem 2 does not hold since $f(x, \theta_0)$ and $f(x, \theta_1)$ do not vanish simultaneously for values of x such that $\theta_0 \leqq x \leqq \theta_1$. Hence, we shall alter our approach to generating $\mathcal{A}$ by proving a theorem which will yield as a consequence a technique for constructing $\mathcal{A}$.

THEOREM 3. *If $(\alpha, \beta, n) \varepsilon A_n$ where $\alpha > 0$ and $n > 0$, then $(\alpha, \beta, n) \not\varepsilon \mathcal{A}$.*

PROOF. If $(\alpha, \beta, n) \varepsilon A_n$, then it follows from (5) that $\beta = (\theta_0/\theta_1)^n(1 - \alpha)$. Consider the test

$$(\alpha', \beta', n') = \alpha(1, 0, 0) + (1 - \alpha)\left(0, \left(\frac{\theta_0}{\theta_1}\right)^n, n\right) = \left(\alpha, (1 - \alpha)\left(\frac{\theta_0}{\theta_1}\right)^n, (1 - \alpha)n\right) = (\alpha, \beta, (1 - \alpha)n).$$

Clearly $(\alpha, \beta, (1 - \alpha)n)$ is preferred to (α, β, n). Theorem 3 states that all single sample tests (α, β, n) such that $0 < \alpha \leqq 1$ and $n > 0$ are inadmissible in the class of mixed single-sample tests. Consequently, the class $\mathcal{A}$ can be generated by the test $(1, 0, 0)$ and the sequence of tests $\{(0, (\theta_0/\theta_1)^k, k)\}$, $k = 0, 1, 2, \cdots$. Since $(\theta_0/\theta_1)^n$ is a convex function of n, it can be shown that $(\alpha, \beta, n) \varepsilon \mathcal{A}$ if and only if $(\alpha, \beta, n) = \gamma_1(1, 0, 0) + \gamma_2(0, \theta_0/\theta_1)^k, k) + \gamma_3(0, (\theta_0/\theta_1)^{k+1}, k + 1$ for some non-negative numbers $\gamma_1, \gamma_2, \gamma_3$ and some non-negative integer k where

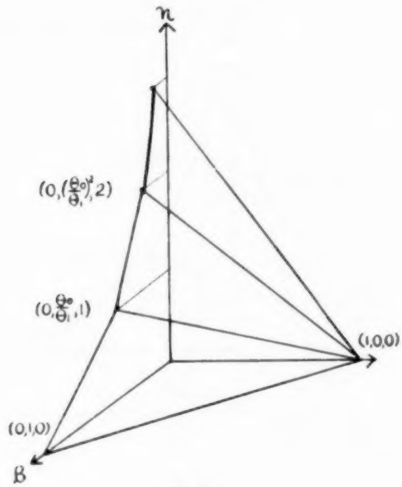


FIG. 7

$\sum_{i=1}^{3} \gamma_i = 1$. In fact it is easily verified that $k = [n/1 - \alpha]$, $\gamma_1 = \alpha$, $\gamma_2 = (1 - \alpha)([n/1 - \alpha] + 1) - n$ and $\gamma_3 = n - (1 - \alpha)[n/1 - \alpha]$. See Fig. 7.

COROLLARY 1. *If $(\alpha, \beta, n) \varepsilon A_n$, there exists a test $(\alpha, \beta, n') \varepsilon \mathfrak{A}$ where $n' = (1 - \alpha)n$.*

PROOF. From the preceding discussion, the test $(\alpha, \beta, n') = \alpha(1, 0, 0) + (1 - \alpha)(0, (\theta_0/\theta_1)^n, n) \varepsilon \mathfrak{A}$. Since $n' = (1 - \alpha)n$, the desired conclusion follows.

We note that the fractional saving in the expected number of observations obtained by randomization is equal to α, *i.e.*,

$$\frac{n - n'}{n} = \frac{n - (1 - \alpha)n}{n} = \alpha.$$

6. Tests on the mean of a rectangular distribution when the range is known. Let

$$\begin{aligned} f(x, \theta) &= 1 && \text{if } \theta < x < \theta + 1, \\ &= 0 && \text{elsewhere.} \end{aligned}$$

We wish to test the hypothesis $H_0: \theta = 0$ against the alternative $H_1: \theta = \theta_1$ where $0 < \theta_1 < 1$. A simple calculation shows that

$$A_n = \{(\alpha, \beta, n): \alpha = (1 - t)^n, \quad \beta = (1 - \theta_1)^n - (1 - t)^n,$$
$$\theta_1 \leqq t \leqq 1\} = \{(\alpha, \beta, n): 0 \leqq \alpha \leqq (1 - \theta_1)^n, \alpha + \beta = (1 - \theta_1)^n\}.$$

See Fig. 8. Let $R_\pi(n) = \min_{(a,b,n)\varepsilon A_n} [\pi\alpha + (1 - \pi)\beta] = \min [\pi(1 - \theta_1)^n, (1 - \pi)(1 - \theta_1)^n] = (1 - \theta_1)^n \min (\pi, 1 - \pi)$. Obviously $R_\pi(n)$ is a convex function of n. It follows that $A_n \subset \mathfrak{A}$. In other words, all fixed sample size tests are admissible in the class of mixed single sample tests.

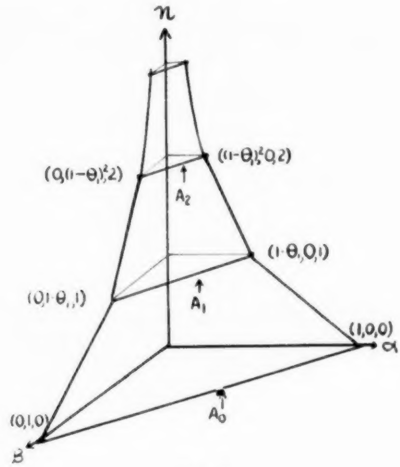


Fig. 8

7. Confidence interval estimation. We next wish to extend the notion of mixed single sample procedures to confidence interval estimation. Perhaps this purpose can best be served by an illustrative example.

Example. Let X denote a normally distributed random variable with unknown mean μ and known variance σ^2. (There is no loss in generality if we assume that $\sigma^2 = 1$, and we shall do so for the remainder of this section.) We wish to consider the problem of obtaining a confidence interval for μ. The standard procedure consists of

(a) choosing a number α between 0 and 1, called the confidence coefficient.
(b) calculating a number t using the equation $\alpha = 1 - 2\Phi(-t)$.
(c) drawing a sample of n independent observations on X and calculating $\bar{X}$, the sample mean.
(d) making the statement tha the interval $(\bar{X} - t/\sqrt{n}, \bar{X} + t/\sqrt{n})$ covers μ with confidence α.

A confidence interval procedure is evaluated in terms of a triple $(1 - \alpha, L, n)$ where $1 - \alpha$ denotes the probability that the confidence interval will not cover μ, L denotes the length of the confidence interval and n denotes the sample size.

We will now exploit the notion of randomizing over the sample size in confidence interval estimation using an approach similar to the one used in Section 2. For integral values of $n \geqq 1$, we let

$$A_n = \left\{(1 - \alpha, L, n) : \alpha = 1 - 2\Phi(t), L = \frac{2t}{\sqrt{n}}, 0 \leqq t < \infty\right\}$$
$$= \left\{(1 - \alpha, L, n) : \alpha = 1 - 2\Phi\left(-\frac{\sqrt{n}L}{2}\right)\right\}.$$

We define $A_0 = \{(1, 0, 0)\}$.

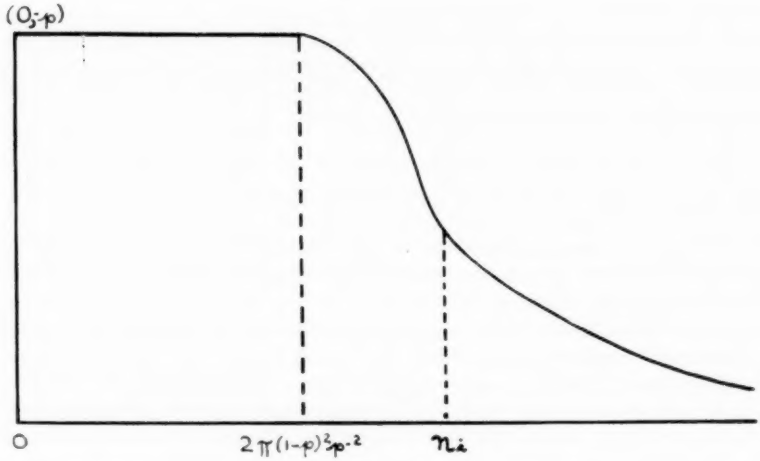


FIG. 9

As an analogue of the Bayes risk $R_\pi(n)$, we consider

$$R_p(n) = \min_{(1-\alpha, L, n)\, \epsilon\, A_n} [p(1-\alpha) + (1-p)L].$$

A routine calculation shows that

$$\begin{aligned} R_p(n) &= p & &\text{if } 0 \leqq n \leqq 2\pi(1-p)^2 p^{-2}, \\ &= 2p\Phi\left(-\sqrt{\log \frac{np^2}{2\pi(1-p)^2}}\right) + \frac{2(1-p)}{n}\sqrt{\log \frac{np^2}{2\pi(1-p)^2}} & & \\ & & &\text{if } n > 2\pi(1-p)^2 p^{-2}, \end{aligned}$$

See Fig. 9.

If we treat $R_p(n)$ as a function of a continuous variable n, we find that

$$R_p'(n) = 0 \qquad \text{if } n < \frac{1}{c},$$

$$R_p''(n) = -\frac{(1-p)}{2} \frac{1 - 3(\log cn)}{\sqrt{n^5}\sqrt{\log cn}} \qquad \text{if } n > \frac{1}{c},$$

where $c = p^2/[2\pi(1-p)^2]$. As in Section 3, there exists a non-negative number $n_i = n_i(p)$ such that $R_p(n)$ is concave on the interval $(0, n_i)$ and convex on the interval (n_i, ∞). In fact, $n_i = [2\pi(1-p)^2]/p^2e^{\frac{1}{3}}$. Using an argument similar to the one used in Section 3, it becomes clear that "improved" mixed confidence interval procedures exist and are of the form

$$(\alpha, L, n) = \gamma(0, 0, 0) + (1-\gamma)(\alpha', L', n'),$$

where $0 < \gamma\ 1$ and (α', L', n') is a fixed sample size confidence interval procedure.

TABLE 3

Confidence coefficient α	Expected sample size n	Length of fixed sample size procedure having the given α and n	Expected length of randomized confidence interval procedure having the given α and n	Percent decrease in the expected length
.044	1	.110	.037	66
.392	9	.334	.329	1
.174	4	.220	.146	34

Table 3 gives some examples of admissible mixed single sample procedures and improvements which can be obtained in the expected length of a confidence interval if a mixing scheme is used.

Improved randomized confidence intervals are of such a nature that certain questions are brought to mind. First, how much "confidence" can we place in randomized confidence intervals? It is true that a confidence interval of the form $(\alpha, L, n) = \gamma(0, 0, 0) + (1 - \gamma)(\alpha', L', n')$ will cover μ 100 α% of the time, will have average length L and will have expected sample size n. However, if we are given confidence interval $(0, 0, 0)$, we no longer have confidence α that we are covering μ. On the other hand, if we are given the confidence interval $(\bar{X} - L'/2, \bar{X} + L'/2)$, we have confidence $\alpha' > \alpha$ that we are covering μ. Furthermore, if a statistician uses a mixed procedure and does not tell this to his customers, then his customers can have confidence α—unless, of course, they are given the procedure $(0, 0, 0)$. (However, if we restrict ourselves to procedures where the sample size n is at least 1, then they could still have confidence α.) In other words, by withholding information from his customers, the statistician gives them confidence α. By giving them information, he either reduces their confidence to 0, or increases their confidence to α'.

This is not the only example of such a situation in statistical techniques. Take, for example, the Stein two sample procedure for finding a confidence interval (of fixed length 1 and confidence coefficient α) for the mean of a normal distribution with unknown variance. A sample of n_0 observations is taken and the sample variance S_0^2 is calculated. Then, an additional n_1 observations are taken where

$$n_1 = \max\left\{n_0, \left[\frac{S_0^2}{d}\right] + 1\right\} - n_0,$$

where d depends on α and 1. The two samples are then combined, the mean $\bar{X}$ of the combined samples is calculated and the confidence interval $\left(\bar{X} - \frac{l}{2}, \bar{X} + \frac{l}{2}\right)$ is given. Now, if it turns out that the variance S^2 of the combined samples is much larger than S_0^2, one is led to believe that the second sample size was not large enough. Thus, one's confidence of α might be reduced, given this information. However, if one did not have this information about S^2, then one's confidence would still be α. This situation is indeed similar to the preceding one.

Another peculiarity of mixed single sample confidence interval procedures is that we get short length only when we do not cover μ. This immediately brings to

mind the question of average length as a criterion for a confidence interval procedure. It is clear that small length is desirable if μ is being covered. What one wants when μ is not covered is open to question. Clearly, we can agree that procedures which give small length when μ is not covered and large length when μ is covered are not desirable ones. Randomized procedures are of this nature.

8. The k decision problem. Let X denote a random variable with distribution function $F(x, \theta)$. Instead of considering only two possible values of θ, θ_0 and θ_1, as we did in the previous section, we now consider k possible values of θ. Let $\theta_1, \theta_2, \cdots, \theta_k$ denote the k possible values of θ. We assume that $\theta_1 < \theta_2 < \cdots < \theta_k$. For any fixed sample size decision rule δ_n, based on samples of size n, let $\alpha_i(\delta_n)$ denote the probability that θ_i will not be selected as the true value of θ when θ_i is the true value of θ if the decision rule δ_n is used. Every fixed sample size decision rule is then identified with an operating characteristic $(\alpha_1, \alpha_2, \cdots \alpha_k, n)$ where $\alpha_i = \alpha_i(\delta_n)$ for $i = 1, 2, \cdots, k$ and where n denotes the sample size. The classes S_n, A_n, A, A^* and $\mathfrak{A}$ are defined in an obvious way and the functions $R_\pi(n)$ and $r_\pi(n)$ are defined as in Section 2 where $\pi = (\pi_1, \pi_2, \cdots \pi_k)$, $\pi_i \geqq 0$ and $\sum_{i=1}^k \pi_i = 1$. We can then extend all the results obtained in Section 2 to the k decision problem.

In the particular case

$$F(x, \theta) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2}\left(\frac{t-\theta}{\sigma}\right)^2\right\} dt,$$

where $\sigma > 0$ is known, we shall show that it is possible to obtain improvements by randomization. For each positive integral value of n, an essentially complete class of decision rules, C_n, can be generated in the following way: Let $(x_1, x_2, \cdots x_n)$ denote a sample of n independent observations on X and let let $(t_0, t_1, \cdots, t_k)$ denote a partition of the real line such that $t_i \leqq t_{i+1}$, $i = 0, 1, \cdots k - 1$. In particular, $t_0 = -\infty$ and $t_k = \infty$. Then any procedure which selects θ_i as the true value of θ whenever $t_{i-1} \leqq \bar{X} < t_i$ is called a monotone procedure. Let C_n denote the class of all monotone procedures. The class C_n is known to be essentially complete.

By definition,

$$\begin{aligned} R_\pi(n) &= \min_{(\alpha_1, \alpha_2, \cdots, \alpha_k) \in A_n} \sum_{i=1}^k \pi_i \alpha_i = \min_{(\alpha_1, \alpha_2, \cdots, \alpha_k) \in C_n} \sum_{i=1}^k \pi_i \alpha_i \\ &= \min_{(t_1, t_2, \cdots, t_{k-1})} \sum_{i=1}^k \pi_i \left[1 - \Phi\left(\sqrt{n}\, \frac{(t_i - \theta_i)}{\sigma}\right) + \Phi\left(\sqrt{n}\, \frac{(t_{i-1} - \theta_i)}{\sigma}\right)\right] \\ &= \sum_{i=1}^k \pi_i \left[1 - \Phi\left(\frac{\xi_i}{\sqrt{n}\delta_i} - \frac{\sqrt{n}\delta_i}{2}\right) + \Phi\left(\frac{\xi_{i-1}}{\sqrt{n}\delta_{i-1}} + \frac{\sqrt{n}\delta_{i-1}}{2}\right)\right], \end{aligned}$$

where

$$\xi_i = \log \frac{\pi_{i+1}}{\pi_i}, \qquad \delta_i = \frac{\theta_i - \theta_{i+1}}{\sigma}, \qquad i = 1, 2, \cdots, k-1,$$

$$\frac{\xi_0}{\sqrt{n}\delta_0} + \frac{\sqrt{n}\delta_0}{2} = -\infty \quad \text{and} \quad \frac{\xi_k}{\sqrt{n}\delta_k} - \frac{\sqrt{n}\delta_k}{2} = \infty.$$

Considering $R_\pi(n)$ as a function of a continuous variable n, we find

$$\text{(a)} \quad R'_\pi(n) = \sum_{i=2}^{k} \pi_i\, \rho\left(\frac{\xi_{i-1}}{\sqrt{n}\delta_{i-1}} + \frac{\sqrt{n}\delta_{i-1}}{2}\right) \cdot \frac{\delta_{i-1}}{\sqrt{n}}.$$

$$\text{(b)} \quad R''_\pi(n) = \sum_{i=2}^{k} \frac{-1}{\delta_n^{5/2}\,\delta_{i-1}}\left(\frac{\xi_{i-1}}{\sqrt{n}\delta_{i-1}} + \frac{\sqrt{n}\delta_{i-1}}{2}\right)(\delta_{i-1}^4 n^2 + 4\delta_{i-1}^2 n - 4\xi_{i-1}^2).$$

For each value of $i = 2, 3, \cdots k$, the function $f_i(n) = \delta_{i-1}^4 n^2 + 4\delta_{i-1}^2 n - 4\xi_{i-1}^2$ is a quadratic function of n. Since the only non-negative root of the equation $f_i(n) = 0$ is

$$n_i = \frac{-2 + 2\sqrt{1 + \xi_{i-1}^2}}{\delta_{i-1}^2},$$

it follows that

$$f_i(n) \leqq 0 \qquad \text{if } 0 \leqq n \leqq n_i\,,$$

$$f_i(n) > 0 \qquad \text{if } n > n_i\,.$$

Then, since $\delta_{i-1} < 0$ for $i = 2, 3, \cdots, k$, it follows that

$$R''_\pi(n) < 0 \qquad \text{if } n < \min_i\,(n_i)$$

and

$$R''_\pi(n) > 0 \qquad \text{if } n > \max_i\,(n_i)$$

Hence, if we let $a = \min_i (n_i)$ and $b = \max_i (n_i)$, it follows that $R_\pi(n)$ is concave on the interval $(0, a)$ and convex on the interval (b, ∞). Clearly, $a \leqq b$.

Thus, for certain values of $\pi_1, \pi_2, \cdots \pi_k$ and $\theta_1, \theta_2, \cdots, \theta_k$, it is possible to achieve improvements by randomization.

9. Testing a composite hypothesis against a composite alternative. We next wish to extend the notion of mixed single sample tests to the problem of testing a composite hypothesis againxt a composite alternative. To fix ideas, let

$$f(x, \theta) = \frac{1}{\sqrt{2\pi}\,\delta} \exp\left\{-\frac{1}{2}\left(\frac{x-\theta}{\sigma}\right)^2\right\},$$

where $\sigma > 0$ is known. We wish to test the hypothesis $H_0{:}\theta \leqq \theta_0$ against the alternative $H_1{:}\theta > \theta_1$, $\theta_1 > \theta_0$. If we are given α and n, the "best" fixed sample size test of level α and size n is obtained by using the best fixed sample size test of $H'_0{:}\theta = \theta_0$ against $H'_1{:}\theta = \theta_1$ corresponding to the given α and n. The resultant fixed sample size test has the desirable property that its power function $P(\theta \mid \alpha, n)$ tends to 1 as θ tends to infinity.

Can we construct, for given α and n, a "good" mixed single sample test of level α and *expected* sample size n in an analogous way? Clearly, if the best mixed single sample test of H'_0 against H'_1 is a bona fide mixture, it is not even true that its power function, $P(\theta)$, approaches 1 as θ approaches infinity. For, in this case,

the fixed sample size test (0, 1, 0) will be chosen with probability λ, say, where $0 < \lambda < 1$, so that $P(\theta) \leqq 1 - \lambda$ for all θ.

However, it should be noted that the fact that $P(\theta)$ does not tend to 1 as θ tends to infinity is not always undesirable for we know, in certain cases, that the set of possible values of θ is bounded, e.g., in testing the mean height θ of American soldiers, we know that $\theta \leqq 6$ feet 2 inches. Consequently, a test procedure which does not have high power at $\theta = 7$ feet is not necessarily undesirable.

Finally, we note that if we restrict ourselves to randomizing over fixed sample size tests of sample size $n > 1$, then $P(\theta) \to 1$ as $\theta \to \infty$.

10. Comparison with the Wald Sequential Probability Ratio Test. In general, it is difficult to compare the improvements attainable by using the Wald Sequential Probability Ratio Test with improvements attainable by randomizing over fixed sample size procedures. For, every test will now be identified with a quadruple $(\alpha, \beta, E_{\theta_0}(n), E_{\theta_1}(n))$. $E_{\theta_0}(n)$ and $E_{\theta_1}(n)$ are usually difficult to calculate. However, in the case of mixed single sample tests, $E_{\theta_0}(n) = E_{\theta_1}(n)$ and do not depend on the unknown value of θ. In some special cases it is easy to make a comparison and this we shall do.

Example.

$$
\begin{aligned}
f(x, \theta) &= \frac{1}{\theta} && \text{if } \theta \leqq x < \theta, \\
&= 0 && \text{elsewhere.}
\end{aligned}
$$

It can be shown that if we use Wald's test, only two types of tests are attainable. They are the test (1, 0, 0, 0) or tests of the form

$$
\left(0, \left(\frac{\theta_0}{\theta_1}\right)^k, k, \frac{1 - \left(\frac{\theta_0}{\theta_1}\right)^k}{1 - \left(\frac{\theta_0}{\theta_1}\right)}\right),
$$

where k is a non-negative integer. However, using mixed single sample tests, we can attain the test (1, 0, 0, 0) and tests of the form $(0, (\theta_0/\theta_1)^k, k, k)$ where k is a non-negative integer, and mixtures of such tests. Since

$$
\lim_{\frac{\theta_0}{\theta_1} \to 1} \frac{k}{\left[\dfrac{1 - \left(\frac{\theta_0}{\theta_1}\right)^k}{1 - \left(\frac{\theta_0}{\theta_1}\right)}\right]} = 1,
$$

it is clear that if θ_0/θ_1 is close to 1, then mixed single sample procedures are almost as good as Wald procedures.

11. Estimation. Can mixing fixed sample estimation procedures yield improvements in estimation techniques? If we evaluate a fixed sample size estimator t_n in terms of a pair of numbers $\{E[L(t_n, \theta)], n\}$, where $E[L(t_n, \theta)]$ denotes the

expected loss if the estimator t_n is used when θ is the true parameter and where n denotes the sample size, then mixing over fixed sample size procedures will not yield improvements since in all problems of practical interest $E[L(t_n, \theta)]$ is a convex function of n. For example, if we wish to estimate the mean θ of a distribution with finite variance σ^2, then, if $t_n = \bar{X}$ and if $L(t_n, \theta) = k(\bar{X} - \theta)^2$, we find that $E[L(t_n, \theta)] = k\,\sigma^2/n$. Thus, it will not pay to randomize.

12. Conclusion. In what situations is a mixed single sample procedure justifiable? In order to answer this question, we must first realize that throughout this paper, we have been judging a test δ by its operating characteristic (α, β, n). If this triple is our only means of evaluating a test procedure, then it is true that single sample procedures would not be justifiable since a sequential probability ratio test achieving the given α and β would be better. However, practical considerations might limit one to a single stage of sampling, e.g., in agricultural experiments, one might not wish to use more than one stage of sampling; or, if one is testing electric light bulbs, one might not wish to test the bulbs sequentially. Other examples could be given.

One could reasonably ask why fixed sample size procedures should not always be used in these situations. Presumably, if the experiment were a so called "one shot affair", i.e., if the experiment were never to be repeated, then one might reasonably insist on a non-randomized fixed sample size procedure (although, of course, this position is not universally held). However, if one repeats the experiment often, it would be reasonable to use a mixed sample size procedure. To illustrate this point, consider Example II in Section 4. In this example, suppose θ_0 represents the probability that a person who has been contaminated with a certain disease will respond positively to a certain test and θ_1 represents the probability that a person who has not been contaminated will respond positively to this same test. Then, if several thousand people are to be classified as either contaminated or non-contaminated according to this test, then the mixed test (1/101, .525, 1) would be preferred to the test (.05, .525, 1) since the mixed test will falsely classify less than 1 percent of the contaminated people whereas the fixed sample size procedure will misclassify 5 percent of the contaminated people. On the other hand, both tests will misclassify the same percentage of non-contaminated people, and both procedures will use on the average of one test per person.

At this point, one could raise strenuous objections to mixed single sample tests on grounds similar to those raised in Section 7, i.e., if one is told which single sample test is actually used, the conditional probabilities of misclassification are no longer α and β. For example, consider a mixed test of the form

$$(\alpha, \beta, n) = \gamma(0, 1, 0) + (1 - \gamma)(\alpha', \beta', n').$$

Now, suppose that a person is told that he has been classified according to the test (0, 1, 0). Such a person would of course be most unhappy. On the other hand, if he is not told which of the tests was used, he would maintain his con-

fidence in the procedure used. In other words, *by withholding information, one can influence a person's willingness to accept a result.* Some feel that axiomatically this is an untenable policy.

13. Acknowledgements. The author is indebted to Allan Birnbaum for suggesting the Bayes approach exploited in this paper. The author is also indebted to Howard Raiffa for his many helpful suggestions and constructive criticisms.

REFERENCES

[1] W. Kruskal, Private communication.

[2] H. Raiffa, "Statistical decision theory approach to item selection for dichotomous test criterion variables," School of Aviation Medicine Series in Statistics, Report No. 56-139, Randolph Field, 1957.

[3] A. Birnbaum, "Applications of vector-valued risk functionals," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 544–545 (abstract).

[4] J. L. Hodges, Jr., and E. L. Lehmann, "The efficiency of some non-parametric competitors of the t-test", *Ann. Math. Stat.*, Vol. 27, No. 2 (1956) pp. 324–335.

ASYMPTOTIC NORMALITY AND EFFICIENCY OF CERTAIN NONPARAMETRIC TEST STATISTICS[1]

By Herman Chernoff and I. Richard Savage

Stanford University and *University of Minnesota*

1. Summary. Let $X_1, \cdots, X_m$ and $Y_1, \cdots, Y_n$ be ordered observations from the absolutely continuous cumulative distribution functions $F(x)$ and $G(x)$ respectively. If $z_{Ni} = 1$ when the ith smallest of $N = m + n$ observations is an X and $z_{Ni} = 0$ otherwise, then many nonparametric test statistics are of the form

$$mT_N = \sum_{i=1}^{N} E_{Ni} z_{Ni} .$$

Theorems of Wald and Wolfowitz, Noether, Hoeffding, Lehmann, Madow, and Dwass have given sufficient conditions for the asymptotic normality of T_N. In this paper we extend some of these results to cover more situations with $F \neq G$. In particular it is shown for all alternative hypotheses that the Fisher-Yates-Terry-Hoeffding c_1-statistic is asymptotically normal and the test for translation based on it is at least as efficient as the t-test.

2. Introduction. Finding the distributions of nonparametric test statistics and establishing optimum properties of these tests for small samples has progressed slower than the corresponding large sample theory. Even so, it is not possible to state that the basic framework of the large sample theory has been completed. Dwass [3] has recently presented a general theorem on the asymptotic normality of certain nonparametric test statistics under alternative hypotheses. His results, however, do not apply to such important and interesting procedures as the c_1-test [11]. Many papers have appeared giving the asymptotic efficiency of particular tests. Hodges and Lehmann [7] have discussed the asymptotic efficiency of the Wilcoxon test with respect to all translation alternatives. In the same paper they have conjectured that the c_1-test is as efficient as the t-test for normal alternatives and at least as efficient as the t-test for all other alternatives.

The beginning of our work came from a desire to verify the Hodges and Lehmann conjecture. Related to the conjecture is the hypothesis that the c_1-statistic is asymptotically normally distributed. Thus our work has two parts: developing a new theorem for asymptotic normality of nonparametric test statistics and the establishing of the variational argument required for determining the minimum efficiency of test procedures.

Received November 26, 1957; revised May 2, 1958.

[1] This work was done with the partial support of the Office of Naval Research while I. R. Savage was at Stanford University and the Center for Advanced Study in the Behavioral Sciences.

Our basic result on the asymptotic normality of statistics of the form T_N is Theorem 1 of Section 4. This theorem is a partial generalization of results of Dwass [3] summarized in our Theorem 4. Theorem 1 is not given in the most general form possible. Our choice of the level of generality was to facilitate our writing and your reading.

Section 3 contains our basic notation and assumptions. Section 4 contains statements of the theorem on asymptotic normality as well as the basic portion of the proof. Details regarding the negligibility of the remainder terms are given in Section 7. The variational arguments are presented in Section 5 and Section 6 relates our Theorem 1 to Dwass's results. Applications of Theorem 1 to several nonparametric tests are given in Section 6.

3. Assumptions and notation. Let $X_1, X_2, \cdots, X_m$ be the ordered observations of a random sample from a population with continuous cumulative distribution function $F(x)$. Let $Y_1, Y_2, \cdots, Y_n$ be the ordered observations of a random sample from a population with continuous cumulative distribution function $G(x)$. Let $N = m + n$ and $\lambda_N = m/N$ and assume that for all N the inequalities $0 < \lambda_0 \leqq \lambda_N \leqq 1 - \lambda_0 < 1$ hold for some fixed $\lambda_0 \leqq \frac{1}{2}$.

Let $F_m(x) =$ (number of $X_i \leqq x)/m$ and $G_n(x) =$ (number of $Y_i \leqq x)/n$. Thus $F_m(x)$ and $G_n(x)$ are the sample cumulative distribution functions of the X's and Y's respectively. Define $H_N(x) = \lambda_N F_m(x) + (1 - \lambda_N)G_n(x)$. Thus $H_N(x)$ is the combined sample cumulative distribution function. The combined population cumulative distribution function is $H(x) = \lambda_N F(x) + (1 - \lambda_N)G(x)$. Even though $H(x)$ depends on N (or rather m and n) through λ_N our notation suppresses this fact for convenience. In fact $F(x)$ and $G(x)$ may actually depend on N although this will not be stated explicitly. In Corollary 1 the distributions do depend on N. The point for suppressing this fact is that our limit theorems are "uniform" and hold, whether the distributions are constant, tend to a limit, or vary rather arbitrarily with the sample size N.

If the ith smallest in the combined sample is an X let $z_{Ni} = 1$ and otherwise let $z_{Ni} = 0$. Then our concern is with statistics of the form

$$mT_N = \sum_{i=1}^{N} E_{Ni}\, z_{Ni}, \tag{3.1}$$

where the E_{Ni} are given numbers. (The special case where $E_{Ni} = E(i/N)$ is particularly easily handled by our methods. For the Wilcoxon test this condition is met with $E_{Ni} = i/N$, and Freund and Ansari [6] have considered $E_{Ni} = E(i/N) = |\frac{1}{2} - i/N|$ in testing for the equality of dispersion of two populations.) The definition (3.1) of T_N is the one conventionally used. We shall, however, use the following representation:

$$T_N = \int_{-\infty}^{\infty} J_N[H_N(x)]\, dF_m(x). \tag{3.2}$$

The definitions (3.1) and (3.2) are equivalent when $E_{Ni} = J_N(i/N)$. A repre-

sentation like (3.2) was used by Blum and Weiss [1, page 243, Eq. 2.4] and R. v. Mises considered $\int \varphi(x)\, dF_m(x)$ in detail [9].

Throughout our proofs K will be used as a generic constant which may depend on J_N but it will not depend on $F(x)$, $G(x)$, m, n, N. Statements involving o_p or O_p will always be uniform in $F(x)$, $G(x)$, and $H(x)$, and λ_N in the interval $0 < \lambda_0 \leq \lambda_N \leq 1 - \lambda_0 < 1$.

While J_N need be defined only at $1/N, 2/N, \cdots, N/N$, we shall find it convenient to extend its domain of definition to $(0,1]$ by some convention such as letting J_N be constant on $(i/N, (i+1)/N]$.

Let I_N be the interval in which $0 < H_N(x) < 1$. Then I_N is closed on the left at the smallest observation and open on the right at the largest observation. The interval, I_N, has a random location.

4. Asymptotic normality.

THEOREM 1. *If*

(1) $J(H) = \lim_{N\to\infty} J_N(H)$ *exists for* $0 < H < 1$ *and is not constant,*

(2) $\displaystyle\int_{I_N} [J_N(H_N) - J(H_N)]\, dF_m(x) = o_p(N^{-1/2})$,

(3) $J_N(1) = o(\sqrt{N})$,

(4) $\displaystyle |J^{(i)}(H)| = \left|\frac{d^iJ}{dH^i}\right| \leq K[H(1-H)]^{-i-\frac{1}{2}+\delta}$

for $i = 0, 1, 2$, *and for some* $\delta > 0$,

then, for fixed F, G *and* λ_N,

$$\lim_{N\to\infty} P\left(\frac{T_N - \mu_N}{\sigma_N} \leq t\right) = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx, \tag{4.1}$$

where

$$\mu_N = \int_{-\infty}^{\infty} J[H(x)]\, dF(x) \tag{4.2}$$

and

$$\begin{aligned} N\sigma_N^2 = 2(1-\lambda_N)\Bigg\{ &\iint_{-\infty<x<y<\infty} G(x)[1-G(y)]J'[H(x)]J'[H(y)]\, dF(x)\, dF(y) \\ &+ \frac{(1-\lambda_N)}{\lambda_N} \iint_{-\infty<x<y<\infty} F(x)[1-F(y)]J'[H(x)]J'[H(y)]\, dG(x)\, dG(y)\Bigg\}, \end{aligned} \tag{4.3}$$

providing $\sigma_N \neq 0$.

In Eqs. 4.1 and 4.3 we put subscripts on μ and σ to recall that these depend on F, G and λ_N and are meaningful in the more general case where F, G, and λ_N are not fixed. Corollary 1 will extend Theorem 1 to obtain convergence to normality uniformly with respect to F, G, and λ_N for a broad range of F, G, and λ_N.

To facilitate the proof of Corollary 1, we will regard F, G, and λ_N as variable throughout the proof of Theorem 1 except where it is specified otherwise.

Assumption 1 is likely to be filled whenever one speaks of a sequence of tests. In the special case $E_{Ni} = E(i/N)$ of course $J_N = E = J$ and Assumption 2 will automatically be satisfied. Theorem 2 shows that Assumptions 1, 2 and 3 are often satisfied when the E_{Ni} are the mean values of order statistics. Assumption 4 is the basic condition. The assumption has two functions: it limits the growth of the coefficients E_{Ni} and it supplies certain smoothness properties. Both conditions are essential to our argument. We believe that the theorem is true without the smoothness condition.

PROOF. To begin the proof we rewrite T_N as

$$T_N = \int_{-\infty}^{\infty} J_N(H_N)\, dF_m(x) = \int_{I_N} [J_N(H_N) - J(H_N)]\, dF_m(x)$$
$$+ \int_{I_N} J(H_N)\, dF_m(x) + \int_{H_N=1} J_N(H_N)\, dF_m .$$

In the second integral we write $dF_m = d(F_m - F + F)$, $J(H_N) = J(H) + (H_N - H)J'(H) + [(H_N - H)^2/2]J''[\varphi H_N + (1 - \varphi)H]$, where $0 < \varphi < 1$, and $H = \lambda_N F + (1 - \lambda_N)G$. After multiplying out the expression becomes

$$T_N = A + B_{1N} + B_{2N} + \sum_{i=1}^{6} C_{iN} ,$$

where

$$A = \int_{0<H<1} J(H)\, dF(x), \tag{4.4}$$

$$B_{1N} = \int_{0<H<1} J(H)\, d[F_m(x) - F(x)], \tag{4.5}$$

$$B_{2N} = \int_{0<H<1} (H_N - H)J'(H)\, dF(x), \tag{4.6}$$

$$C_{1N} = \lambda_N \int_{0<H<1} (F_m - F)J'(H)\, d[F_m(x) - F(x)] \tag{4.7}$$

$$C_{2N} = (1 - \lambda_N) \int_{0<H<1} (G_n - G)J'(H)\, d[F_m(x) - F(x)], \tag{4.8}$$

$$C_{3N} = \int_{I_N} \frac{(H_N - H)^2}{2} J''[\varphi H_N + (1 - \varphi)H]\, dF_m(x), \tag{4.9}$$

$$C_{4N} = \int_{H_N=1} [- J(H) - (H_N - H)J'(H)]\, dF_m(x), \tag{4.10}$$

$$C_{5N} = \int_{I_N} [J_N(H_N) - J(H_N)]\, dF_m(x), \tag{4.11}$$

$$C_{6N} = \int_{H_N=1} J_N(H_N)\, dF_m(x). \tag{4.12}$$

The A, B, C terms represent the "constant," "first order random," and "higher order random" portions respectively of T_N. In this section a detailed study of the A and B terms is made and in Section 7 it is shown that the C terms are of higher order.

The "constant" term, $A = \int_{0<H<1} J(H)\, dF(x)$, is finite as a result of Assumption 4 of Theorem 1; see Section 7.A.10. Since A depends on λ_N as well as $F(x)$ and $G(x)$ it need not converge as $N \to \infty$, but it does remain bounded.

Integrating B_{2N} by parts and using the fact that

$$\int_{-\infty}^{\infty} d[F_m(x) - F(x)] = 0,$$

we obtain

$$B_{1N} + B_{2N} = [1 - \lambda_N]\left\{\int_{-\infty}^{\infty} B(x)\, d[F_m(x) - F(x)] - \int_{-\infty}^{\infty} B^*(x)\, d[G_n(x) - G(x)]\right\}, \tag{4.13}$$

where

$$B(x) = \int_{x_0}^{x} J'[H(y)]\, dG(y) \tag{4.14}$$

$$B^*(x) = \int_{x_0}^{x} J'[H(y)]\, dF(y) \tag{4.15}$$

and

$$\lambda_N B^*(x) + (1 - \lambda_N)B(x) = J[H(x)] - J[H(x_0)]$$

with x_0 determined somewhat arbitrarily, say by $H(x_0) = 1/2$.

Thus,

$$B_{1N} + B_{2N} = [1 - \lambda_N]\left\{\frac{1}{m}\sum_{i=1}^{m} [(BX_i) - \mathcal{E}B(X)] - \frac{1}{n}\sum_{i=1}^{n} [B^*(Y_i) - \mathcal{E}B^*(Y)]\right\}, \tag{4.16}$$

where $\mathcal{E}$ represents expectation and X and Y have the F and G distributions respectively.

The two summations involve independent samples of identically distributed random variables. Therefore, if F, G, and λ_N are fixed, $B(X)$ and $B^*(Y)$ are specified random variables and we may apply the central limit theorem to show that $B_{1N} + B_{2N}$ when properly normalized has a Gaussian distribution in the limit. The central limit theorem applies if the variances of $B(X)$ and $B^*(Y)$ are finite and at least one is positive.

First, we shall find a bound on the moments of $B(X)$ and $B^*(Y)$:

$$| B(x) | = \left| \int_{x_0}^{x} J'[H(y)]\, dG(y) \right| \leqq K[H(x)[1 - H(x)]]^{-\frac{1}{2}+\delta}.$$

Thus for $\delta' > 0$ such that $(2 + \delta')(-\frac{1}{2} + \delta) > -1$,

$$\mathcal{E}\{|B(X)|\}^{2+\delta'} \leq K \int_{-\infty}^{\infty} [H(x)[1 - H(x)]]^{(-\frac{1}{2}+\delta)(2+\delta')}\, dF(x)$$

$$\leq K \int_0^1 [H(1 - H)]^{(-\frac{1}{2}+\delta)(2+\delta')}\, dH \leq K,$$

having made use of $dG \leq (1/\lambda_0)\, dH$. (See Section 7.A.8.)

Similarly, we may bound the $2 + \delta'$ absolute moments of $B^*(Y)$. The asymptotic normality of $B_{1N} + B_{2N}$ follows providing $B(X)$ and $B^*(Y)$ do not both have zero variance.

We compute the variances of $B(X)$ and $B^*(Y)$. These can be expressed in terms of $\int B(x)\, dF(x)$, $\int B^2(x)\, dF(x)$, etc., but we shall use a slightly different approach.

$$B(X) - \mathcal{E}B(X) = \int_{-\infty}^{\infty} B(x)\, d[F_1(x) - F(x)]$$

$$= -\int_{-\infty}^{\infty} [F_1(x) - F(x)]J'[H(x)]\, dG(x)$$

has variance

$$\sigma^2_{B(X)} = \mathcal{E}\left\{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [F_1(x) - F(x)][F_1(y) - F(y)]J'[H(x)]J'[H(y)]\, dG(x)\, dG(y)\right\},$$

and

$$\sigma^2_{B(X)} = 2 \iint_{-\infty<x<y<\infty} F(x)[1 - F(y)]J'[H(x)]J'[H(y)]\, dG(x)\, dG(y), \tag{4.17}$$

if it is permitted to interchange expectation and integral. That this may be done follows from Fubini's theorem when it is seen that for $x < y$,

$$\mathcal{E}\{|F_1(x) - F(x)|\,|F_1(y) - F(y)|\} \leq KF(x)[1 - F(y)]$$

and that the last integral above is finite. (In fact this integral is bounded in the argument dealing with (C_{23N}) in Section 7.B.)

Similarly, the variance of $B^*(Y)$ is given by

$$\sigma^2_{B^*(Y)} = 2 \iint_{-\infty<x<y<\infty} G(x)[1 - G(y)]J'[H(x)]J'[H(y)]\, dF(x)\, dF(y). \tag{4.18}$$

These two variances when combined give the variance result stated in (4.3). We review the status of our proof. In Section 7, the C terms are shown to be "higher order uniformly." The A term is non-random and finite. Finally

$$B_{1N} + B_{2N}$$

is the sum of two independent terms each of which is the average of random variables with mean 0 and finite second moments. Theorem 1 follows.

The proof given can be extended to the case where F, G and λ_N are not fixed. To obtain uniform convergence to normality, we apply a theorem of Esseen

([4], p. 43) which is a generalization of the so-called Berry-Esseen theorem ([8], p. 288)[2]. Since the C terms are uniformly $o_p(1/\sqrt{N})$ it suffices to obtain uniform convergence for $B_{1N} + B_{2N}$. For this it suffices to bound $\rho_{2+\delta'}$ for $B(X)$ and $B^*(Y)$. Since we bounded the absolute $2 + \delta'$ moments, all that is required is to bound the variances of $B(X)$ and $B^*(Y)$ away from 0 and to have m and $n \to \infty$. Thus we have

Corollary 1. *If the conditions* 1 *to* 4 *of Theorem* 1 *are satisfied, and* F, G, *and* $\lambda_N(0 < \lambda_0 \leqq \lambda_N \leqq 1 - \lambda_0 < 1)$ *are restricted to a set for which* $B(X)$ *and* $B^*(Y)$ *have variances bounded away from* 0, *then Eq.* 4.1 *(asymptotic normality) holds uniformly with respect to* F, G, *and* λ_N.

Corollary 2. *If conditions* 1 *to* 4 *of Theorem* 1 *are satisfied,* $0 < \lambda_0 \leqq \lambda_N \leqq 1 - \lambda_0 < 1$,

$$F(x) = \Psi(x - \theta_N),$$
$$G(x) = \Psi(x - \varphi_N),$$

where Ψ *has a density* ψ, *then Eq.* 4.1 *holds uniformly with respect to* λ_N, θ_N *and* φ_N *for* $\varphi_N - \theta_N$ *in some neighborhood of* 0. *If* $\varphi_N - \theta_N \to 0$,

$$\lim_{N\to\infty} \frac{\lambda_N N\sigma_N^2}{(1 - \lambda_N)} = 2 \iint\limits_{0<x<y<1} x(1 - y)J'(x)J'(y)\,dx\,dy = \int_0^1 J^2(x)\,dx - \left[\int_0^1 J(x)\,dx\right]^2. \tag{4.19}$$

Proof. It suffices to show that $B(X)$ and $B^*(Y)$ have variances bounded away from zero and to establish Eq. 4.19. Since J is not constant and has a second derivative, there is an interval of u in which $J'(u)$ is bounded away from 0 and in which $J'(u) > 0$ or in which $J'(u) < 0$. There is a corresponding interval of x for which $\Psi(x)$ lies in the u interval and its density $\psi(x)$ is almost everywhere bounded away from 0. For $\varphi_N - \theta_N$ small enough, there is an x interval whose length is bounded away from 0 where the densities $f(x) = \psi(x - \theta_N)$ and $g = \psi(x - \varphi_N)$ are almost everywhere bounded away from 0 and $J'[H(x)]$ is bounded away from zero. It follows that $B(X)$ and $B^*(Y)$ have variances bounded away from zero.

All that remains is to establish Eq. 4.19. The first equality follows directly from Theorem 1 by letting $F(x) = x^*$ and $G(x) \to x^*$. The second equality can be obtained by interpreting the double integral as

$$\iiiint\limits_{0<u<x<y<v<1} J'(x)J'(y)\,du\,dx\,dy\,dv$$

[2] Esseen's theorem states that if $X_1, X_2, \cdots, X_n$ are independent observations from a population with mean 0, variance σ^2, and finite absolute $2 + \delta'$ moment $\beta_{2+\delta'}$, $0 < \delta' \leqq 1$, then $|F^* - \Phi^*| < C(\delta')\left[\frac{\rho_{2+\delta'}}{n^{\delta'/2}} + \frac{\rho_{2+\delta'}^{1/\delta'}}{n^{1/2}}\right]$ where F^* is the cdf. of $\bar{X}$, Φ^* is the approximating normal cdf, C depends only on δ' and $\rho_{2+\delta'} = \beta_{2+\delta'}/\sigma^{2+\delta'}$.

and integrating with respect to y first and x second. It can also be obtained by considering a standard derivation [13] of the asymptotic distribution of T_N when $F = G$ where T_N is regarded as the average of a sample of m from the population of N numbers $J_N(1/N), J_N(2/N), \cdots, J_N(N/N)$.

We remark that normalizing J so that $\int_0^1 J(x)\,dx = 0$ and $\int_0^1 J^2(x)\,dx = 1$ will not affect the efficiency of the test. Furthermore, if J is the inverse of a cdf, the right-hand side of (4.19) is the variance of that distribution.

In applying Theorem 1 the verification of condition 2 may cause some difficulty. The following Theorem 2 gives a simple sufficient condition under which conditions 1, 2, and 3 hold. In particular with the use of Theorem 2 it is simple to verify that the distribution of the c_1-statistic does approach a Gaussion distribution for alternative hypotheses.

THEOREM 2. *If $J_N(i/N)$ is the expectation of the ith order statistic of a sample of size N from a population whose cumulative distribution function is the inverse function of J and*

$$|J^{(i)}(u)| \leqq K[u(1-u)]^{-i-\frac{1}{2}+\delta}, \qquad i = 0, 1, 2,$$

then

$$\lim_{N\to\infty} J_N(H) = J(H), \qquad 0 < H < 1,$$

$$J_N(1) = o(N^{1/2}),$$

and

$$\int_{I_N} [J_N(H_N) - J(H_N)]\,dF_m(x) = o(N^{-1/2}).$$

(We write o instead of o_p because the random sequence is bounded by a nonrandom sequence which is $o(N^{-1/2})$. In fact $|\int [J_N(H_N) - J(H_N)]\,dF_m(x)| \leqq (1/\lambda)\int |J_N(H_N) - J(H_N)|\,dH_N(x)$ and our proof essentially shows that this latter integral which is non-random and independent of F and G, is $o(N^{-1/2})$.)

PROOF. It is well known that $J_N(H) \to J(H)$. A proof of the other two results is given in Section 7.C.

5. Variational argument. We have now established that the limiting distribution of the c_1-statistic is Gaussian. Thus we may proceed with the study of the efficiency of this test procedure. We will examine translation alternatives only. Since the power of the c_1-test approaches one when the distributions F and G are held fixed as N approaches infinity we restrict our consideration to the following situation.

There is a distribution function $\Psi(x)$ which does not depend on N and $F(x) = \Psi(x - \theta)$ and $G(x) = \Psi(x - \varphi)$. We test the hypothesis that $\Delta = \theta - \varphi = 0$ vs. "near" alternatives of the form $\Delta = \Delta_N = cN^{-1/2}$. We will also assume that

$$0 < \lim_{N\to\infty} \lambda_N = \lambda < 1.$$

With this framework we are able to use the Pitman criterion (the one considered

by Hodges and Lehmann) for finding efficiencies of test procedures. The following conditions have been established for the c_1-statistic if Ψ has a density and clearly hold for the t-statistic if Ψ has finite second moments. There are functions $a_N(\Delta)$ and $b_N(\Delta)$ such that for Δ in some neighborhood of 0,

$$\mathcal{L}\left(\frac{T_N - a_N(\Delta)}{b_N(\Delta)}\right) \Rightarrow N(0, 1), \tag{5.1}$$

$$\operatorname*{limit}_{N\to\infty} \frac{b_N(\Delta_N)}{b_N(0)} = 1, \tag{5.2}$$

$$E_T = \operatorname*{limit}_{N\to\infty} \left[\frac{[a_N(\Delta_N) - a_N(0)]}{\Delta_N N^{1/2} b_N(0)}\right]^2 \tag{5.3}$$

exists and is independent of c.

The quantity E_T is called the efficacy of the procedure based on the sequence of statistics T_N. Of course E_T depends on Ψ. In comparing two sequences of tests, say T_N and T_N^*, for the same pair of near alternatives the two tests will have the same power only when the corresponding sample sizes, N and N^*, satisfy the following relationship

$$\operatorname*{limit}_{N\to\infty} \frac{N^*}{N} = \frac{E_T}{E_{T^*}} = E_{T,T^*} \tag{5.4}$$

if $E_{T^*} \neq 0$. *E_{T,T^*} is called the asymptotic relative efficiency of T_N with respect to T_N^*.*

Let $E_{c_1,t}(\Psi)$ denote the asymptotic efficiency relative to the t-test of the c_1-test against translation alternatives. Then we have $J = J_0$ the inverse of the normal $N(0, 1)$ cdf Φ and applying Corollary 1 and using derivatives in the expression for E_T, we have

$$E_{c_1,t}(\Psi) = I_{1\Psi}^2/\sigma^2, \tag{5.5}$$

where

$$I_{1\Psi} = \int J_0'[\Psi(x)]\psi^2(x)\, dx \tag{5.6}$$

and σ^2 is the variance of the distribution with cdf Ψ (and density ψ).[3] Normalizing Ψ to have mean 0 and variance 1 does not affect $E_{c_1,t}(\Psi)$ which then becomes equal to $I_{1\Psi}^2$. In this section we shall prove

THEOREM 3. *If Ψ is a cdf with a density and finite second moment, then $E_{c_1,t}(\Psi) \geqq 1$, and $E_{c_1,t}(\Psi) = 1$ only if Ψ is normal.*

PROOF. It suffices to show that the minimum of $I_{1\Psi}$ subject to the restrictions

$$I_{2\Psi} = \int x\psi(x)\, dx = 0$$

[3] If Ψ does not have finite variance σ^2, $E_{c_1,t}$ is not defined but it makes sense to regard it as ∞.

and

$$I_{3\Psi} = \int x^2\psi(x)\,dx = 1$$

is attained only for $\Psi = \Phi$ and that $I_{1\Phi} = 1$.

A density $\psi(x)$ assigns to each x a value of Ψ and a corresponding value of $J_0[\Psi(x)]$. If $\psi(x) = 0$ a.e. on an interval, this interval corresponds to a fixed value of $J_0[\Psi(x)]$. If x is then regarded as a function of J_0, it is multivalued at that value of J_0. Otherwise x is continuous and it is increasing in J_0. Conversely any monotone non-decreasing function x of J_0 determines a corresponding cdf Ψ. We have

$$u = \Phi[J_0(u)],$$

$$J_0'(u) = \frac{1}{\varphi[J_0(u)]},$$

and

$$\varphi(t) = \frac{1}{\sqrt{2\pi}}\,e^{-t^2/2}.$$

Furthermore

$$\int_{-\infty}^{x}\psi(t)\,dt = \Psi(x) = \int_{-\infty}^{J_0}\varphi(t)\,dt \tag{5.7}$$

and

$$\psi(x)\,dx = d\Psi(x) = \varphi(J_0)\,dJ_0\,.$$

Consequently our problem consists of finding a monotone function $x(J_0)$ which minimizes

$$I_{1\Psi} = \int \frac{1}{\varphi(J_0)}\frac{\varphi(J_0)}{\left(\frac{dx}{dJ_0}\right)}\varphi(J_0)\,dJ_0 = \int \frac{\varphi(J_0)}{\left(\frac{dx}{dJ_0}\right)}\,dJ_0 \tag{5.8}$$

subject to the restrictions and

$$I_{2\Psi} = \int x\psi(x)\,dx = \int x\varphi(J_0)\,dJ_0 = 0, \tag{5.9}$$

$$I_{3\Psi} = \int x^2\psi(x)\,dx = \int x^2\varphi(J_0)\,dJ_0 = 1. \tag{5.10}$$

In the above form it is immediately obvious that if $\Psi = \Phi$, $x = J_0$ and hence $I_{1\Phi} = 1$. This form is also more suitable for our variational approach.

Suppose now that x is replaced by $x^* = cx$. Then I_1, I_2 and I_3 are replaced by $I_1^* = I_1/c$, $I_2^* = cI_2$, and $I_3^* = c^2I_3$. Thus if $I_2 = 0$ and $I_3 < 1$, we can obtain $I_2^* = 0$ and $I_3^* = 1$ with $I_1^* < I_1$. This discussion is relevant to the proof of the following lemma.

LEMMA 1. *The solution of the minimization problem is unique if it exists.*

PROOF. Suppose x_1 and x_2 are distinct functions with non-negative derivatives. Then let $x = (1 - w)x_1 + wx_2$, where $0 \leqq w \leqq 1$. Then, by convexity

$$I_1(w) = \int \frac{\varphi(J_0)}{\left(\frac{dx}{dJ_0}\right)} dJ_0 < (1 - w) \int \frac{\varphi(J_0)}{\left(\frac{dx_1}{dJ_0}\right)} dJ_0 + w \int \frac{\varphi(J_0)}{\left(\frac{dx_2}{dJ_0}\right)} dJ_0,$$

$$I_2(w) = \int x\varphi(J_0)\, dJ_0 = (1 - w) \int x_1\varphi(J_0)\, dJ_0 + w \int x_2\varphi(J_0)\, dJ_0,$$

and

$$I_3(w) = \int x^2\varphi(J_0)\, dJ_0 < (1 - w) \int x_1^2\varphi(J_0)\, dJ_0 + w \int x_2^2\varphi(J_0)\, dJ_0.$$

Hence x_1 and x_2 cannot both be solutions of the minimization problem since otherwise a multiple of $(x_1 + x_2)/2$ would then satisfy the side conditions and yield a smaller I_1.

With this lemma, all that remains is to show that $x = J_0$ is a solution of the problem. To this end we establish a sufficient condition for the solution of the problem as follows. Suppose that x_1 and x_2 are monotone functions satisfying the restrictions where x_2 gives a lower value for I_1 than does x_1. Then using the convexity again, we have

$$I_1'(0) = -\int \frac{\frac{d(x_2 - x_1)}{dJ_0}}{\left(\frac{dx_1}{dJ_0}\right)^2} \varphi(J_0)\, dJ_0 < 0,$$

$$I_2'(0) = \int (x_2 - x_1)\varphi(J_0)\, dJ_0 = 0,$$

and

$$I_3'(0) = 2 \int x_1(x_2 - x_1)\varphi(J_0)\, dJ_0 < 0.$$

Consequently we have

LEMMA 2. *If x_1 satisfies the restrictions and if for each x_2 which does so also there is a $\xi \geqq 0$ such that*

$$I_1'(0) + \xi I_3'(0) \geqq 0,$$

then x_1 is the unique solution of the minimization problem.[4]

[4] This sufficient condition is essentially the usual Euler equation except that with the convexity at our disposal and the monotonicity restriction, it plays the role of a sufficient instead of a necessary condition.

Now

$$I_1'(0) = \frac{-(x_2 - x_1)}{\left(\frac{dx_1}{dJ_0}\right)^2}\varphi(J_0)\Big|_{-\infty}^{\infty} + \int (x_2 - x_1)\left[\frac{\varphi'(J_0)}{\left(\frac{dx_1}{dJ_0}\right)^2} - \frac{\frac{2d^2x_1}{dJ_0^2}\varphi(J_0)}{\left(\frac{dx_1}{dJ_0}\right)^3}\right] dJ_0 .$$

Now let $x_1(J_0) = J_0$. Then

$$I_1'(0) + \xi I_3'(0) = \int (x_2 - x_1)[\varphi'(J_0) + 2\xi J_0 \varphi(J_0)]\, dJ_0 ,$$

which vanishes for $\xi = 1/2$. Applying Lemma 2 establishes our theorem.

If we regarded the c_1-test as one tailor made to compete against the best parametric test for translation when F and G are normal, we may inquire about nonparametric tests designed to compete against the best parametric tests when F and G have some other form.

Suppose F and G are known to be of the form $F_0(x - \theta)$ and $F_0(x - \varphi)$ respectively where F_0 has a twice differentiable density f_0. Then an efficient[5] test statistic for $\Delta = \theta - \varphi = 0$ would be the maximum-likelihood estimate

$$\hat{\Delta} = \hat{\theta} - \hat{\varphi}$$

for which the asymptotic distribution is normal with mean Δ and variance $[N\lambda(1 - \lambda)(\mathrm{inf}_{F_0})]^{-1}$, where

$$\mathrm{inf}_{F_0} = \int \frac{[f_0'(x)]^2}{f_0(x)}\, dx, \tag{5.11}$$

providing the above integral exists. The relative efficiency of our nonparametric test based on the test statistic T with a specified normalized[6] J to the $\hat{\Delta}$ test is

$$E_{T,\hat{\Delta}}(F_0) = \frac{I_{1F_0}^2}{\mathrm{inf}_{F_0}}, \tag{5.5a}$$

where

$$I_{1F_0} = \int J'(F_0) f_0^2(x)\, dx. \tag{5.6a}$$

It can be shown that the best J in the sense that it maximizes $E_{T,\hat{\Delta}}(F_0)$ is given by

$$J(u) = \frac{-f_0'(x)}{f_0(x)}\, (\mathrm{inf}_{F_0})^{-1/2} \tag{5.12}$$

[5] There seems to be no clear-cut statement in the literature which would establish the test based on $\hat{\Delta}$ as an efficient test invariant under the same translation of the X_i and Y_j. The authors wish to thank the referee who pointed out the following elegant proof. The efficacy of the $\hat{\theta} - \hat{\varphi}$ test is $\lambda(1 - \lambda)\,\mathrm{inf}_{F_0}$, where inf_{F_0} is the information of F_0. No invariant test of $\Delta = \Delta_N$ vs. $\Delta = 0$ can have greater efficacy than the likelihood ratio test for testing $\Delta = \Delta_N$ vs. $\Delta = 0$ when the densities of X and Y are $f_0(x + (1 - \lambda)\Delta)$ and $f_0(x - \lambda\Delta)$. A standard calculation gives this test efficacy $\lambda(1 - \lambda)\,\mathrm{inf}_{F_0}$. Thus our $\hat{\theta} - \hat{\varphi}$ test is efficient.

[6] Let J be normalized so that $\int J(u)\, du = 0$ and $\int J^2(u)\, du = 1$.

where $u = F_0(x)$. In fact for this J, we have

$$I_{1F_0} = -(\inf_{F_0})^{-1/2} \int \left[\frac{f_0''(x)}{f_0(x)} - \frac{[f_0'(x)]^2}{f_0^2(x)}\right] \frac{1}{f_0(x)} f_0'(x)\, dx = (\inf_{F_0})^{1/2}$$

and

$$E_{T,\hat{\Delta}} = 1.$$

As it is to be expected, if $F_0 = \Phi(N(0, 1))$, the corresponding $J = J_0$, the inverse of Φ. The problem of comparing the nonparametric with the parametric procedures designed for F_0 when F and G are translates of $\Psi \neq F_0$ is hindered by our ignorance of the behavior of the *parametric* procedure when $\Psi \neq F_0$.

6. Orientation and applications.

6.A. *Orientation.* In Fraser's book [5] it is shown that the c_1-test has a limiting normal distribution for normal alternatives. We have now shown this to be the case for all alternatives (if we include the cases where $N\sigma_N^2 = 0$ or $N\sigma_N^2 \to 0$ as degenerate cases). Hoeffding's U-statistics include many nonparametric test statistics and he, Lehmann, and Dwass have shown that U-statistics are asymptotically normal under the alternative hypothesis. The U-statistics do not include all statistics of the form

$$mT_N = \sum_{i=1}^{N} E_{Ni} z_{Ni}. \tag{3.1}$$

In particular c_1 is not a U-statistic. Dwass's results [3], summarized in Theorem 4, appear to be the only useful results for statistics of the form (3.1) under general alternative hypotheses.

THEOREM 4. *Suppose*

(1) *The conditions of the first paragraph of our Section 3 hold (Dwass has written to us indicating that it is sufficient to have m and n approach ∞);*

(2) *The polynomial*

$$P(t) = \sum_{k=1}^{h} b_k t^k$$

is non-degenerate, i.e.,

$$\max(|b_1|, \cdots, |b_h|) > 0;$$

(3) $(X_1, \cdots, X_m, Y_1, \cdots, Y_n) = (U_1, \cdots, U_N)$ *and R_i is the number of U's less than or equal to U_i,*

$$a_{Ni} = \begin{cases} a_1 = (n/mN)^{1/2}, & i = 1, \cdots, m, \\ a_2 = -(m/nN)^{1/2}, & i = m+1, \cdots, N; \end{cases} \tag{4}$$

$$t_N = \sum_{i=1}^{N} a_{Ni} P(R_i/N); \tag{5}$$

then

$$\lim_{N\to\infty} P\left(\frac{t_N - E(t_N)}{\sigma_{t_N}} < s\right) = \int_{-\infty}^{s} \frac{e^{-x^2/2}}{\sqrt{2\pi}}\, dx.$$

First note

$$\begin{aligned} t_N &= \sum_{i=1}^{N} P\left(\frac{i}{N}\right)[a_1 z_{Ni} + a_2(1 - z_{Ni})] \\ &= \sum_{i=1}^{N} P\left(\frac{i}{N}\right) z_{Ni} \left(\frac{1}{\sqrt{N}}\left[\left(\frac{1-\lambda_N}{\lambda_N}\right)^{1/2} + \left(\frac{\lambda_N}{1-\lambda_N}\right)^{1/2}\right]\right) + a_2 \sum_{i=1}^{N} P\left(\frac{i}{N}\right) \\ &= \sqrt{N}\, T_N \left(\left(\frac{1-\lambda_N}{\lambda_N}\right)^{1/2} + \left(\frac{\lambda_N}{1-\lambda_N}\right)^{1/2}\right) + K, \end{aligned}$$

where in T_N we have $E_{Ni} = P(i/N)$. Thus there is a non-stochastic linear relationship between t_N and T_N. Hence, from the statistical viewpoint t_N is equivalent to T_N, a statistic of the form (3.1). Now let us compare Dwass's conditions with ours.

(1) Requiring λ_N to be bounded away from 0 and 1 seems to be essential in our Theorem 1.

(2) The condition $E_{Ni} = J_N(i/N) = P(i/N) = \sum_{k=1}^{h} b_k(i/N)^k$ is much stronger than our condition 4 in Theorem 1 in two respects: We only require that $J_N(x)$ have a limit and the limit need not be a polynomial in x. Of particular importance we do not require $J(x)$ to be bounded on $0 < x < 1$. The requirement $\max(|b_1|, \cdots, |b_h|) > 0$ is to insure that $E_{Ni} \equiv 0$, a trivial case which causes no difficulty.

6.B. *Applications.*

Example 1: Let $E_{Ni} = \sum_{j=i}^{N} j^{-1}$. Then Savage has proved [10] that T_N has a limiting Gaussian distribution under the hypothesis and is the test statistic for the locally most powerful rank test of $\theta_1 = \theta_2$ against the alternative $\theta_1 \neq \theta_2$ where $F(x) = e^{\theta_1 x}$ and $G(x) = e^{\theta_2 x}$, $-\infty < x \leq 0$ and $F(x) = G(x) = 1$, $x > 0$. In order to verify that T_N has a limiting Gaussian distribution under the alternative hypothesis let us check the conditions of Theorem 1. To do so we note that $J_N(i/N)$ is the expected value of the ith smallest observation of a sample from the exponential distribution and that Theorem 2 is applicable. Hence T_N is asymptotically normal in all cases.

Example 2: Van der Waerden [12] has developed the theory of the test statistic

$$T_N = \int_{-\infty}^{\infty} J\left(\frac{NH_N(x)}{N+1}\right) dF_m(x),$$

where J is the inverse of the normal $N(0, 1)$ cumulative distribution. It can be shown that

$$\int_{-\infty}^{\infty} \left| J\left(\frac{NH_N(x)}{N+1}\right) - J(H_N(x)) \right| dH_N(x) = o\left(\frac{1}{\sqrt{N}}\right).$$

Then conditions 2 and 3 of Theorem 1 are established and the asymptotic nor-

mality and efficiency properties for this statistic are verified to be the same as those of the c_1-statistic.

7. Higher order terms. In proving that the C terms of Theorem 1 are uniformly of higher order the following elementary results are used repeatedly.

7.A. *Elementary results.*

1. $H \geqq \lambda_N F \geqq \lambda_0 F$.
2. $H \geqq (1 - \lambda_N)G \geqq \lambda_0 G$.
3. $1 - F \leqq \dfrac{1 - H}{\lambda_N} \leqq \dfrac{1 - H}{\lambda_0}$.
4. $1 - G \leqq \dfrac{1 - H}{1 - \lambda_N} \leqq \dfrac{1 - H}{\lambda_0}$.
5. $F(1 - F) \leqq \dfrac{H(1 - H)}{\lambda_N^2} \leqq \dfrac{H(1 - H)}{\lambda_0^2}$.
6. $G(1 - G) \leqq \dfrac{H(1 - H)}{\lambda_0^2}$.
7. $dH \geqq \lambda_N \, dF \geqq \lambda_0 \, dF$.
8. $dH \geqq (1 - \lambda_N) \, dG \geqq \lambda_0 \, dG$.
9. Let (a_N, b_N) be the interval $S_{N\epsilon}$, where

$$S_{N\epsilon} = \left\{ x : H(1 - H) > \frac{\eta_\epsilon \lambda_0}{N} \right\}. \tag{7.1}$$

Then η_ϵ can be chosen independently of F, G and λ_N so that

$$P\{X_i \in S_{N\epsilon}, Y_j \in S_{N\epsilon}, i = 1, 2, \cdots, m, j = 1, 2, \cdots, n\} \geqq 1 - \epsilon. \tag{7.2}$$

10. $\int_{-\infty}^{\infty} J(H(x)) \, dF(x)$ is finite.

PROOF. Using assumption 4 of Theorem 1 and A.7

$$\begin{aligned} \left| \int_{-\infty}^{\infty} J(H(x)) \, dF(x) \right| &\leqq K \int_0^1 [H(1 - H)]^{-1+\delta} \, dH \\ &\leqq K \int_0^1 \frac{dH}{[H(1 - H)]^{1/2}} \leqq K. \end{aligned} \tag{7.3}$$

7.B. *Detailed consideration of the second order terms of Theorem* 1. We are now ready to show that the C terms are uniformly of higher order. We begin with C_{1N} and prove the following identity:

$$\begin{aligned} C_{1N} &= \lambda_N \int_{-\infty}^{\infty} (F_m - F) J'(H) \, d(F_m(x) - F(x)) \\ &= \frac{\lambda_N}{2} \left[\int J'(H) \, d(F_m - F)^2 + \frac{1}{m} \int J'(H) \, dF_m \right]. \end{aligned} \tag{7.4}$$

Let R be the set of points of increase of F_m. Then the right-hand side of the identity becomes

$$\frac{\lambda_N}{2}\left[\int_{\bar{R}} J'(H)\,d(F_m - F)^2 + \int_R J'(H)\,d(F_m - F)^2 + \frac{1}{m}\sum_{i=1}^m J'(H(X_i))\,\frac{1}{m}\right]$$

$$= \frac{\lambda_N}{2}\left[2\int_{\bar{R}} J'(H)(F_m - F)\,d(F_m - F) + \sum_{i=1}^m J'(H(X_i))\right.$$

$$\left.\cdot\left[\left(\frac{i}{m} - F(X_i)\right)^2 - \left(\frac{i-1}{m} - F(X_i)\right)^2\right] + \frac{1}{m}\sum_{i=1}^m J'(H(X_i))\,\frac{1}{m}\right]$$

$$= \frac{\lambda_N}{2}\left[2\int_{\bar{R}} J'(H)(F_m - F)\,d(F_m - F) + \sum_{i=1}^m J'(H(X_i))\right.$$

$$\left.\cdot\left[\frac{2}{m}\left[\frac{i}{m} - F(X_i)\right] - \frac{1}{m^2}\right] + \sum_{i=1}^m J'(H(X_i))\,\frac{1}{m^2}\right]$$

$$= \lambda_N\int (F_m - F)J'(H)\,d(F_m - F).$$

Using this identity we integrate by parts and obtain

$$C_{1N} = -\frac{\lambda_N}{2}\,(C_{11N} + C_{12N} - C_{13N}), \tag{7.5}$$

where

$$C_{11N} = \int_{S_{N\epsilon}} (F_m - F)^2 J''(H)\,dH,$$

$$C_{12N} = \int_{\bar{S}_{N\epsilon}} (F_m - F)^2 J''(H)\,dH,$$

$$C_{13N} = \frac{1}{m}\int J'(H(x))\,dF_m(x)$$

$$= \frac{1}{m^2}\sum_{i=1}^m J'(H(X_i)),$$

where $S_{N\epsilon}$ was defined in 7.A.9.

Now let us consider the random variable C_{11N}. We find

$$\mathcal{E}\,|\,C_{11N}\,| = \mathcal{E}\left\{\int_{S_{N\epsilon}} (F_m - F)^2\,|\,J''(H)\,|\,dH\right\} = \int_{S_{N\epsilon}} \frac{F(1 - F)}{N\lambda_N}\,|\,J''(H)\,|\,dH.$$

Now using assumption 4 of Theorem 1 and 7.A.5 we obtain

$$\mathcal{E}\,|\,C_{11N}\,| \leqq \frac{K}{N}\int_{S_{N\epsilon}} \frac{H(1 - H)\,dH}{[H(1 - H)]^{\frac{5}{2}-\delta}}$$

$$\leqq \frac{K}{N}\int_{\frac{1}{KN}}^1 \frac{1}{H^{\frac{3}{2}-\delta}}\,dH$$

$$\leqq \frac{K}{N^{\frac{1}{2}+\delta}}.$$

Now using the Markoff inequality ([2], p. 182),

$$\Pr\left(|C_{11N}| > aN^{-1/2}\right) \leqq \frac{K}{N^{\frac{1}{2}+\delta}} \frac{N^{1/2}}{a} = \frac{K}{aN^{\delta}},$$

where K may depend on ϵ. Now consider C_{12N}.

Let $H_1 = H(a_N)$, $H_2 = H(b_N)$ as in 7.A.9. Then $H_1 = 1 - H_2 < K/N$. With probability greater than $1 - \epsilon$ we have

$$C_{12N} = \int_{S_{N\epsilon}} (F_m - F)^2 J''(H)\, dH = \int_0^{H_1} F^2 J''(H)\, dH + \int_{H_2}^1 (1 - F)^2 J''(H)\, dH$$

$$|C_{12N}| \leqq K\left[\int_0^{H_1} \frac{H^2\, dH}{(H(1 - H))^{\frac{5}{2}-\delta}} + \int_{H_2}^1 \frac{(1 - H)^2\, dH}{(H(1 - H))^{\frac{5}{2}-\delta}}\right]$$

$$\leqq K \int_0^{H_1} H^{-\frac{1}{2}+\delta}\, dH \leqq KN^{-\frac{1}{2}-\delta}.$$

Hence $C_{11N} + C_{12N}$ which does not involve ϵ is $o_p(N^{-\frac{1}{2}})$. Now to complete the study of C_{1N} we investigate C_{13N}:

$$|C_{13N}| = \frac{1}{m^2}\left|\sum_{i=1}^m J'[H(X_i)]\right| \leqq \frac{K}{m^2} \sum_{i=1}^m [H(X_i)(1 - H(X_i))]^{-\frac{3}{2}+\delta}.$$

We may assume $\delta < \frac{3}{2}$ or $\delta < \frac{1}{2}$ without loss of generality. Then using 7.A.5

$$|C_{13N}| \leqq \frac{K}{N}\frac{1}{m} \sum_{i=1}^m [F(X_i)[1 - F(X_i)]]^{-\frac{3}{2}+\delta},$$

which is distribution free. By a theorem of Marcinkiewicz ([8], pp. 242–243) if a random variable Y has rth order moment finite $(0 < r < 1)$, then the sum of N independent observations on Y is $o_p(N^{1/r})$. If X has cdf F,

$$[F(X)[1 - F(X)]]^{-\frac{3}{2}+\delta}$$

has a finite moment of order $2/(3 - \delta)$ and hence

$$C_{13N} = o_p\left[\frac{1}{m^2} N^{\frac{3}{2}-\frac{\delta}{2}}\right] = o_p[N^{-\frac{1}{2}}].$$

Consequently $C_{1N} = o_p(N^{-\frac{1}{2}})$.

Next consider

$$C_{2N} = (1 - \lambda_N) \int_{-\infty}^{\infty} (G_n - G)J'(H)\, d[F_m(x) - F(x)]. \tag{7.6}$$

We have

$$C_{2N} = (1 - \lambda_N)(C_{21N} + C_{22N})$$

where

$$C_{21N} = \int_{S_{N\epsilon}} (G_n - G)J'(H)\, d[F_m(x) - F(x)],$$

$$C_{22N} = \int_{S_{N\epsilon}} (G_n - G)J'(H)\, d[F_m(x) - F(x)].$$

With probability greater than $1 - \epsilon$, there are no observations in $\bar{S}_{N\epsilon}$ and

$$|C_{21N}| \leqq K \int_{S_{N\epsilon}} H(1 - H)[H(1 - H)]^{-\frac{3}{2}+\delta}\, dH(x) \leqq K \left(\frac{\eta_\epsilon}{N}\right)^{\frac{1}{2}+\delta}.$$

Since the two samples are independent and $\mathcal{E}(G_n - G) = 0$, we have

$$\mathcal{E}(C_{22N}) = \mathcal{E}\{\mathcal{E}C_{22N} \mid X_1, X_2, \cdots, X_m\} = 0,$$

$$\mathcal{E}(C_{22N}^2 \mid X_1, X_2, \cdots, X_m) = C_{23N} + C_{24N},$$

$$C_{23N} = \frac{2}{n} \iint\limits_{\substack{x,y \in S_{N\epsilon} \\ x<y}} G(x)[1 - G(y)]J'[H(x)]J'[H(y)]$$

$$\cdot d[F_m(x) - F(x)]\, d[F_m(y) - F(y)],$$

$$C_{24N} = \frac{1}{nm} \int_{S_{N\epsilon}} G(x)[1 - G(x)]\{J'[H(x)]\}^2\, dF_m(x),$$

$$\mathcal{E}(C_{23N}) = \frac{-2}{nm} \iint\limits_{\substack{x,y \in S_{N\epsilon} \\ x<y}} G(x)[1 - G(y)]J'[H(x)]J'[H(y)]\, dF(x)\, dF(y)^7$$

$$\leqq \frac{K}{N^2} \iint\limits_{x<y} H(x)[1 - H(y)]\, |J'[H(x)]J'[H(y)]|\, dH(x)\, dH(y)$$

$$\leqq \frac{K}{N^2} \iint\limits_{0<x<y<1} x^{-\frac{1}{2}+\delta}(1 - x)^{-\frac{1}{2}+\delta} y^{-\frac{1}{2}+\delta}(1 - y)^{-\frac{1}{2}+\delta}\, dx\, dy \leqq \frac{K}{N^2},$$

$$\mathcal{E}(C_{24N}) = \frac{1}{nm} \int_{S_{N\epsilon}} G(1 - G)(J'[H])^2\, dF(x) \leqq \frac{K}{N^2} \int_{S_{N\epsilon}} [H(1 - H)]^{-2+2\delta}\, dH(x)$$

$$\leqq \frac{K\eta_\epsilon^{-1+2\delta}}{N^{1+2\delta}} = o(N^{-1}).$$

Hence

$$\mathcal{E}(C_{22N}^2 \mid X_1, X_2, \cdots, X_m) \leqq K o_p(N^{-1}),$$

where K may depend on ϵ and

$$|C_{22N}| \leqq K o_p(N^{-1/2})$$

since

$$P(C_{22N}^2 > a\mathcal{E}(C_{2N}^2 \mid X_1, \cdots, X_m)) < 1/a.$$

Consequently $C_{2N} = (1 - \lambda_N)(C_{21N} + C_{22N})$ which does not involve ϵ, satisfies

$$C_{2N} = o_p(N^{-1/2}).$$

---

[7] This integrand has already appeared as part of the variance in Eq. (4.3).

Now consider

$$C_{3N} = \int_{0<H_N(x)<1} [H_N(x) - H(x)]^2 J''[\varphi H_N(x) + (1-\varphi)H(x)]\, dF_m(x), \tag{7.7}$$

$$0 < \varphi < 1.$$

With probability greater than $1 - \epsilon$, the range of integration $0 < H_N(x) < 1$ can be replaced by $S_{N,\epsilon}$ without changing C_{3N}. Since

$$\sup_{H_N>0} \left| \frac{H(x)}{H_N(x)} \right| = O_p(1), \tag{7.8a}$$

and

$$\sup_{H_N<1} \left| \frac{1-H(x)}{1-H_N(x)} \right| = O_p(1), \tag{7.8b}$$

for each $\epsilon > 0$, there is an $\eta_\epsilon^* > 0$ such that with probability greater than $1 - \epsilon$, we have for $0 < H_N(x) < 1$,

$$[\varphi H_N + (1-\varphi)H][1 - (\varphi H_N + (1-\varphi)H)] > \eta_\epsilon^* H(x)[1 - H(x)]. \tag{7.9}$$

Then

$$| C_{3N} | \leqq \int_{S_{N,\epsilon}} [H_N(x) - H(x)]^2 (\eta_\epsilon^*)^{-1+\delta} \{H[1-H]\}^{-1+\delta}\, dF_m(x) = (\eta_\epsilon^*)^{-1+\delta} C_{31N},$$

$$\begin{aligned} \mathcal{E}(| C_{31N} |) \leqq \frac{1}{N} \int_{S_{N,\epsilon}} \Big[\lambda_N F(1-F) + \frac{(1-F)(1-2F)}{N} \\ + (1-\lambda_N)G(1-G) \Big] [H(1-H)]^{-1+\delta}\, dF(x) \\ \leqq \frac{K}{N} \int_{S_{N,\epsilon}} [H(1-H)]^{-1+\delta}\, dH + \frac{K}{N^2} \int_{S_{N,\epsilon}} [H(1-H)]^{-1+\delta}\, dF \\ \leqq \frac{K\eta_\epsilon^{-\frac{1}{2}+\delta}}{N^{\frac{1}{2}+\delta}} + \frac{K\eta_\epsilon^{-\frac{1}{2}+\delta}}{N^{\frac{1}{2}+\delta}}. \end{aligned}$$

Consequently

$$C_{3N} = o_p(N^{-1/2}).$$

The C_{4N} term vanishes unless the greatest of the $N = m + n$ observations is an X. In that case

$$C_{4N} = \frac{1}{m} \{ -J[H(X_m)] - [1 - H(X_m)]J'[H(X_m)]\}. \tag{7.10}$$

Using 7.A.9, however,

$$\frac{1}{m} | J[H(X_m)] | \leqq \frac{[H(X_m)[1 - H(X_m)]]^{-\frac{1}{2}+\delta}}{m} \leqq \frac{(\eta_\epsilon^*)^{-\frac{1}{2}+\delta}}{N^{\frac{1}{2}+\delta}}$$

with probability at least $1 - \epsilon$. Hence

$$\frac{1}{m} J[H(X_m)] = o_p(N^{-1/2}).$$

Similarly

$$\left| \frac{[1 - H(X_m)]J'[H(X_m)]}{m} \right| \leqq \frac{[1 - H(X_m)]}{m} \{H(X_m)[1 - H(X_m)]\}^{-\frac{3}{2}+\delta}$$

$$\leqq \frac{\{[H(X_m)][1 - H(X_m)]\}^{-\frac{1}{2}+\delta}}{mH(X_m)} = o_p(N^{-1/2})O_p(1) = o_p(N^{-1/2}).$$

Hence

$$C_{4N} = o_p(N^{-1/2}).$$

The neglibility of C_{5N} and C_{6N} follows immediately from Assumptions 2 and 3 of Theorem 1.

7.C. *Proof of Theorem* 2. First we note that

$$J_N\left(\frac{i}{N}\right) = E_{N,i} = \int_0^1 J(u)g_{i,N}(u)\,du, \tag{7.11}$$

where

$$g_{i,N}(u) = \frac{N!}{(i-1)!\,(N-i)!} u^{i-1}(1-u)^{(N-i)} \tag{7.12}$$

is the density of the ith order statistic from the uniform distribution on $[0, 1]$ and incidentally has mean $i/(N+1)$ and variance $i(N-i+1)/[(N+1)^2(N+2)]$. Then we have

$$\begin{aligned} | E_{N,1} | &\leqq KN \int_0^1 [u(1-u)]^{-\frac{3}{2}+\delta}(1-u)^{N-1}\,du \\ &= \frac{KN\Gamma(N - \frac{1}{2} + \delta)\Gamma(\frac{1}{2} + \delta)}{\Gamma(N + 2\delta)} \leqq KN^{\frac{1}{2}-\delta}. \end{aligned} \tag{7.13}$$

By a symmetric argument the desired result $J_N(1) = o(N^{1/2})$ follows. Furthermore we have

$$\left| J_N\left(\frac{1}{N}\right) - J\left(\frac{1}{N}\right) \right| \leqq KN^{\frac{1}{2}-\delta} + K\left[\frac{1}{N}\left(1 - \frac{1}{N}\right)\right]^{-\frac{1}{2}+\delta} \leqq KN^{\frac{1}{2}-\delta}. \tag{7.14}$$

Before proceeding to bound $J_N(i/N) - J(i/N)$ for $1 < i \leqq N/2$ we apply the Stirling formula

$$\begin{aligned} \log x! &= \log \Gamma(x+1) \\ &= \frac{1}{2}\log 2\pi - x + \left(x + \frac{1}{2}\right)\log x + \frac{\theta}{12x}, \qquad 0 < \theta < 1, \end{aligned} \tag{7.15}$$

with a rather standard argument to obtain for $1 < i \leqq N/2, 0 < u \leqq (i-1)/(N-1)$,

$$g_{i,N}(u) \leqq \sqrt{\frac{(N-1)^3}{2\pi(i-1)(N-i)}}\, e^{-\frac{v^2}{2}\left[\frac{(N-1)}{(i-1)(N-i)}\right]}\left[1+\frac{K}{N}\right], \tag{7.16}$$

(7.16)[5]

where

$$v = (N-1)u - (i-1). \tag{7.17}$$

For $1 < i \leqq N/2$,

$$\begin{aligned} J_N\left(\frac{i}{N}\right) - J\left(\frac{i}{N}\right) &= \int_0^1 \left[J(u) - J\left(\frac{i}{N}\right)\right] g_{i,N}(u)\, du \\ &= D_{11} + D_{12} + D_{21} + D_{22} + D_3 + D_4, \end{aligned} \tag{7.18}$$

where

$$D_{11} = \int_0^{u_1} J(u) g_{i,N}(u)\, du, \qquad D_{12} = \int_{1-u_1}^1 J(u) g_{i,N}(u)\, du,$$

$$D_{21} = -\int_0^{u_1} J\left(\frac{i}{N}\right) g_{i,N}(u)\, du, \qquad D_{22} = -\int_{1-u_1}^1 J\left(\frac{i}{N}\right) g_{i,N}(u)\, du,$$

$$D_3 = \int_{u_1}^{1-u_1} \left(u - \frac{i}{N}\right) J'\left(\frac{i}{N}\right) g_{i,N}(u)\, du$$

$$D_4 = \frac{1}{2}\int_{u_1}^{1-u_1} \left(u - \frac{i}{N}\right)^2 J''(u^*) g_{i,N}(u)\, du,$$

u^* between u and i/N, and $u_1 = (i-1)/[2(N-1)]$.

$$\begin{aligned} g_{i,N}(u) &= u^\alpha \frac{u^{i-1-\alpha}(1-u)^{N-i}(N-\alpha)!}{(i-1-\alpha)!\,(N-i)!}\, \frac{N!}{(N-\alpha)!}\, \frac{(i-1-\alpha)!}{(i-1)!} \\ &\leqq K u^\alpha N^\alpha g_{i-\alpha, N-\alpha}(u), \end{aligned} \tag{7.19}$$

where $\alpha = \frac{1}{2} - \delta$ and we assume $\delta < \frac{1}{2}$ and thus $\alpha > 0$ without loss of generality. Let Φ be the normal cdf. Then

$$\begin{aligned} |D_{11}| \leqq \int_0^{u_1} K[u(1-u)]^{-\alpha} K u^\alpha N^\alpha g_{i-\alpha, N-\alpha}(u)\, du \leqq K N^\alpha \\ \cdot \Phi\left[\frac{\left(u_1 - \frac{i-1-\alpha}{N-1-\alpha}\right)(N-1-\alpha)^{3/2}}{\sqrt{(i-1-\alpha)(N-i)}}\right], \end{aligned}$$

$$|D_{11}| \leqq K N^\alpha \Phi\left(\frac{-\sqrt{i}}{K}\right). \tag{7.20}$$

[5] K represents a generic constant independent of i, N, λ_N, F, and G. This equation is related to the asymptotic normality of order statistics and is derived by an operation similar to the direct proof of the asymptotic normality of the binomial distribution.

Since $g_{i,N}(u) \geqq g_{i,N}(1-u)$ for $1 < i \leqq N/2$ and $0 \leqq u \leqq 1/2$, $|D_{12}|$ has the same bound as $|D_{11}|$. Similarly

$$(7.21) \qquad |D_{21}| \leqq K\left(\frac{i}{N}\right)^{-\alpha}\Phi\left[\frac{\left(u_1 - \frac{i-1}{N-1}\right)(N-1)^{3/2}}{\sqrt{(i-1)(N-i)}}\right] \leqq KN^{\alpha}\Phi\left(\frac{-\sqrt{i}}{K}\right)$$

and $|D_{22}|$ has the same bound too. Since the expectation of the ith order statistic from the uniform distribution is $i/(N+1)$,

$$D_3 = -J'\left(\frac{i}{N}\right)\left\{\int_0^{u_1}\left(u - \frac{i}{N}\right)g_{i,N}(u)\,du + \int_{1-u_1}^{1}\left(u - \frac{i}{N}\right)g_{i,N}(u)\,du + \frac{i}{N(N+1)}\right\}$$

Now

$$h(u) = \left|u - \frac{i}{N}\right| g_{i,N}(u) \leqq Kh(1-u) \qquad \text{for } u < u_1.$$

Hence

$$(7.22) \qquad \begin{aligned} |D_3| &\leqq K\left(\frac{i}{N}\right)^{-\alpha-1}\left[K\frac{i}{N}\Phi\left(\frac{-\sqrt{i}}{K}\right) + \frac{i}{N(N+1)}\right] \\ &\leqq KN^{\alpha}\Phi\left(\frac{-\sqrt{i}}{K}\right) + KN^{\alpha-1} \end{aligned}$$

Finally

$$|D_4| \leqq Ku_1^{-\frac{1}{2}+\delta}\int_0^1\left(u - \frac{i}{N}\right)^2 g_{i,N}(u)\,du,$$

$$|D_4| \leqq Ku_1^{-\frac{1}{2}+\delta}\left[\frac{i(N-i+1)}{(N+1)^2(N+2)} + \left(\frac{i}{N+1} - \frac{i}{N}\right)^2\right],$$

$$(7.23) \qquad |D_4| \leqq Ku_1^{-\frac{1}{2}+\delta}\left[K\frac{u_1}{N} + K\frac{u_1^2}{N^2}\right] \leqq \frac{Ku_1^{-\frac{1}{2}+\delta}}{N} \leqq \frac{KN^{\alpha}}{i^{1+\alpha}}.$$

Thus, for $1 < i \leqq N/2$,

$$(7.24) \qquad J_N\left[\frac{i}{N}\right] - J\left[\frac{i}{N}\right] \leqq KN^{\alpha}\left[\Phi\left(\frac{-\sqrt{i}}{K}\right) + \frac{1}{N} + \frac{1}{i^{1+\alpha}}\right]$$

and

$$(7.25) \qquad \begin{aligned} &\left|\int_{1 \leqq NF_m \leqq N/2}[J_N(H_N) - J(H_N)]\,dF_m\right| \\ &\qquad \leqq \frac{1}{m}\left\{KN^{\frac{1}{2}-\delta} + \sum_{i=2}^{N/2} KN^{\alpha}\left[\Phi\left(\frac{-\sqrt{i}}{K}\right) + \frac{1}{N} + \frac{1}{i^{1+\alpha}}\right]\right\} \\ &\qquad \leqq KN^{-\frac{1}{2}-\delta} \end{aligned}$$

since $\sum_{i=1}^{\infty} \Phi(-\sqrt{i}/K)$ and $\sum_{i=1}^{\infty} i^{-(1+a)}$ converge. By a symmetric argument we can cover the range $N/2 < NF_m \leqq N$ and our theorem follows.

BIBLIOGRAPHY

[1] J. R. BLUM AND LIONEL WEISS, "Consistency of certain two-sample tests," *Ann. Math. Stat.*, Vol. 28 (1957), pp. 242–246.

[2] HARALD CRAMÉR, *Mathematical Methods of Statistics*, Princeton, 1946.

[3] MEYER DWASS, "The large sample powers of rank order tests in the two sample problem," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 352–374.

[4] C. G. ESSEEN, "Fourier analysis of distribution functions," *Acta Math.*, Vol. 77 (1945), pp. 1–125.

[5] D. A. S. FRASER, *Nonparametric Methods of Statistics*, John Wiley and Sons, Inc., 1956.

[6] J. E. FREUND AND A. R. ANSARI, "Two-way rank sum test for variances," Virginia Polytechnic Institute report supported by the Office of Ordnance Research and National Science Foundation, 1957.

[7] J. L. HODGES, JR., AND ERICH L. LEHMANN, "The efficiency of some nonparametric competitors of the t-test," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 324–335.

[8] MICHEL LOÈVE, *Probability Theory*, D. Van Nostrand Co., Inc., 1955.

[9] R. V. MISES, "On the asymptotic distribution of differentiable statistical functions," *Ann. Math. Stat.*, Vol. 18 (1947), pp. 309–348.

[10] I. R. SAVAGE, "Contributions to the theory of rank order statistics: Two sample case," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 590–616.

[11] M. E. TERRY, "Some rank order tests which are most powerful against specific parametric alternatives," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 346–366.

[12] B. L. VAN DER WAERDEN, "The Computation of the X-distribution," *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, University of California Press, 1956, pp. 207–208.

[13] ABRAHAM WALD AND JACOB WOLFOWITZ, "Statistical tests based on permutations of the observations," *Ann. Math. Stat.*, Vol. 15 (1944), pp. 358–372.

A HIGH DIMENSIONAL TWO SAMPLE SIGNIFICANCE TEST[1]

By A. P. Dempster

Bell Telephone Laboratories[2], Murray Hill, New Jersey

0. Summary. The classical multivariate 2 sample significance test based on Hotelling's T^2 is undefined when the number k of variables exceeds the number of within sample degrees of freedom available for estimation of variances and covariances. Addition of an a priori Euclidean metric to the affine k-space assumed by the classical method leads to an alternative approach to the same problem. A test statistic F which is the ratio of 2 mean square distances is proposed and 3 methods of attaching a significance level to F are described. The third method is considered in detail and leads to a "non-exact" significance test where the null hypothesis distribution of F depends, in approximation, on a single unknown parameter r for which an estimate must be substituted. Approximate distribution theory leads to 2 independent estimates of r based on nearly sufficient statistics and these may be combined to yield a single estimate. A test of F nominally at the 5% level but based on an estimate of r rather than r itself has a true significance level which is a function of r. This function is investigated and shown to be quite near 5%. The sensitivity of the test to a parameter measuring statistical distance between population means is discussed and it is shown that arbitrarily small differences in each individual variable can result in a detectable overall difference provided the number of variables (or, more precisely, r) can be made sufficiently large. This sensitivity discussion has stated implications for the a priori choice of metric in k-space. Finally a geometrical description of the case of large r is presented.

1. Introduction. The statistical problem here treated is that of significance testing for the difference of the means of 2 k-variate populations which may be assumed to have the same structure of variances and covariances, the test being based on a sample from each population with sample sizes denoted by n_1 and n_2. It is intended to provide a method applicable to data where the number k of characteristics measured on each individual is large but where the number of individuals measured may be quite small. The usual method of classical multivariate statistics encounters a mathematical barrier and becomes inapplicable when $k > n_1 + n_2 - 2$, but certainly the need has arisen in applied statistical work for techniques handling small samples of highly described individuals.

The classical method has 2 equivalent formulations in terms of the T^2 statistic of Hotelling [2] or the best linear discriminator of Fisher [3]. For this method the

Received July 8, 1957; revised June 27, 1958.

[1] Most of the material presented here is also contained in the author's PhD thesis [1] at Princeton University. His work in Princeton was supported principally by the National Research Council of Canada.

[2] Now at Harvard University.

space of the k characteristics is thought of as k-dimensional affine space and needs no further structure: the method is invariant over the choice of any k linear combinations of full rank of the given k variables to be used in place of the given variables. The 2 populations are assumed to be probability distributions over affine k-space and the samples constitute $n_1 + n_2$ points of this space. In the formulation of [3] the sample points are projected along a family of parallel $(k - 1)$-dimensional hyperplanes onto any line, the family being chosen so that the one-dimensional Student's t for the 2 samples is maximized. This $t_{\max}$ is then used to test the significance of the difference in population means. However, if $k > n_1 + n_2 - 2$, a family of $(k - 1)$-dimensional hyperplanes can be chosen which projects the points into 2 points, one for each sample. Then $t_{\max} = \infty$ regardless of the populations and so is useless as a test-statistic. In the formulation of [2] the samples are used to define a Euclidean metric in the affine k-space and the test-statistic is the distance between the 2 sample means in this metric. This metric is based on the variation of the samples about their means, and if the samples are shifted to have a common mean point and $k > n_1 + n_2 - 2$ the variation spans only a subspace of $n_1 + n_2 - 2$ dimensions. Thus it is not surprising that in this case the method of defining the metric breaks down. Furthermore it is heuristically evident that no metric for a whole affine space can be well-defined from variation taking place in a flat subspace. For these reasons we are forced to give up the classical approach with its elegant mathematical property of affine invariance.

The approach of this paper is based on the observation that, whatever metric is chosen for k-space, the distance between sample means is a statistic which may yield evidence of separation of the populations, and, rather than be preoccupied with a choice of optimum metric from the data, we should try to use a metric determined apart from the data and analyze the information yielded through this metric.

For much of the theory the population distributions will be assumed to be (multivariate) normal.

2. The general method. It is assumed that a Euclidean metric has been assigned to the affine k-space of the k characteristics; that is, k independent linear combinations of the given variables have been chosen which define distance along k mutually orthogonal axes of Euclidean k-space. The metric may be thought of as chosen from a priori knowledge (precise or imprecise) of the joint distributions of the k characteristics, in the hope of roughly sphericalizing these distributions. More detailed remarks on the choice of a metric are to be found in section 5.

Suppose that the 2 population distributions have means denoted by $k \times 1$ vectors ν_1 and ν_2 and common $k \times k$ matrix of variances and covariances denoted by Λ. We are seeking evidence that $\nu_0 = \nu_1 - \nu_2$ is different from zero and are naturally led to consider V_0 the vector joining the sample means. V_0 is an unbiased estimate of ν_0. Having a metric at hand we will try to direct a significance test at the detection of a non-zero length of ν_0 and will use the

length of V_0 in estimating this length. Rejection of the null hypothesis $\nu_0 = 0$ will result from evidence that the length of V_0 is significantly greater than zero.

So far this use of V_0 has been justified mostly on heuristic grounds. It makes sense geometrically. If however we assume that the populations are multivariate normal $N(\nu_1, \Lambda)$ and $N(\nu_2, \Lambda)$ a more mathematical reason may be given. Suppose the $n_1 + n_2$ individuals are regarded as defining a set of orthogonal axes in a Euclidean space of $n_1 + n_2$ dimensions. The space may be regarded as "degree of freedom" (d.f.) space and any set of orthogonal axes defines a set of orthogonal d.f. Such a new set of d.f. may be defined as follows: first choose the d.f. measuring the grand mean of the $n_1 + n_2$ individuals, second choose the d.f. measuring the difference between the means of the 2 samples, and third choose any set of $n_1 + n_2 - 2$ d.f. which together with the first 2 form an orthogonal set. This last set represents "within sample" d.f. Their number $n_1 + n_2 - 2$ will henceforth for convenience be denoted by m. The data, which consists of k points in this $(n_1 + n_2)$-space, can be described by a set of $n_1 + n_2$ $k \times 1$ vectors corresponding to the new d.f. Let U_0 be the vector corresponding to the mean difference d.f. and $U_1, U_2, \cdots, U_m$ be the vectors corresponding to the within sample d.f. It can be easily checked that

$$V_0 = \left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{\frac{1}{2}} U_0,$$

that $U_1, U_2, \cdots, U_m$ have mean 0, and that $U_0, U_1, \cdots, U_m$ are uncorrelated and each have Λ for matrix of variances and covariances. Finally, assuming normality and defining

$$\xi = \left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{-\frac{1}{2}} \nu_0,$$

it is seen that $U_0, U_1, \cdots, U_m$ are independent, the first being distributed as $N(\xi, \Lambda)$ and the remainder as $N(0, \Lambda)$. With the normality assumption it is clear that $U_0, U_1, \cdots, U_m$ are sufficient for the parameters ν_0 and Λ, for apart from an irrelevant overall translation of both samples the original data can be reconstructed. But since U_0 is the only one of these vectors involving the parameter ν_0 it is natural to choose a property of U_0 alone in testing significance.

Three methods of testing whether or not U_0 is significantly long will be described, but only the third of these will be pursued. The first is the non-parametric randomization test based on the method of Pitman and Welch [4, 5]. For each of the $\binom{n_1 + n_2}{n_1}$ divisions of the $n_1 + n_2$ individuals into 2 groups of n_1 and n_2 there is a corresponding d.f. for group difference and corresponding vector U. Under the null hypothesis that the $n_1 + n_2$ individuals are a sample from one distribution the lengths of all these vectors U have a joint distribution symmetric under permutation of the vectors. Accordingly U_0 is significantly long at level α if the length of U_0 is beyond the $(1 - \alpha)$ point of the sample cumulative distribution of the set of $\binom{n_1 + n_2}{n_1}$ lengths of vectors U. The second method is

the continuous analogue of the first method which comes into play when normal distributions are assumed. Suppose a set of p d.f. are chosen independently at random uniformly with regard to direction in that part of $(n_1 + n_2)$-space orthogonal to the d.f. for the grand mean. The set of p corresponding vectors together with U_0 have, under the null hypothesis of identical normally distributed populations, joint distributions which are again symmetric under permutations so that a significance test may be defined as in the first method. The limiting test as $p \to \infty$ is uniquely defined and may be regarded as the continuous analogue of the Pitman and Welch procedure. For $k = 1$ this amounts to the usual t test, but for general k the distribution associated with the limiting test appears difficult to handle analytically. However the test could be approximated using a suitable p and experimental sampling.

The third method, which is the concern of most of the subsequent discussion, is also based on normal distribution theory. The idea here is to compare the length of U_0 directly against the lengths of $U_1, U_2, \cdots, U_m$, since under the null hypothesis they form a sample of size $m + 1$ from a certain distribution. Define Q_i = squared length of $U_i (i = 0, 1, \cdots, m)$ and

$$F = Q_0 \bigg/ \frac{1}{m} \sum_{i=1}^{m} Q_i$$

Then U_0 will be declared significantly long if F is significantly large. If the null hypothesis distribution of F involved no unknown parameters then an exact test could be based on F; since this is not the case a type of "non-exact significance test" will be introduced.

3. Distribution theory. The distributions involved in the non-exact significance test are those of properties of the vectors $U_0, U_1, U_2, \cdots, U_m$, in particular their lengths and angles between pairs of them. We suppose in this section normal distributions and so may deal with a typical vector U distributed as $N(0, \Lambda)$ or a typical sample of such vectors. Under these assumptions Q, the squared length of U, has the distribution of a quadratic form in k normal variables. Since this distribution in precise form involves k parameters, all unknown, we will rely on the well-known [6] approximation which treats Q as distributed as $\mu\chi_r^2$ depending only on 2 unknown parameters μ and r. The parameters μ and r are generally fitted by equating the first 2 moments, and this results in the inequality $r \leqq k$.

This approximation, at least for integral r, corresponds to approximating the distribution of vector U by a spherical normal distribution lying in a flat subspace of dimension r in k-space. Stated more precisely this says that in the metric chosen for k-space there is an orthogonal transformation to coordinates $(y_1, y_2, \cdots, y_k)$ such that the distribution of U is defined by

(i) density $\dfrac{1}{(2\pi)^{r/2}} \exp\left(-\dfrac{1}{2\mu}\sum_{i=1}^{r} y_i^2\right)$ for $y_1, y_2, \cdots y_r$, and

(ii) $y_{r+1}, y_{r+2}, \cdots, y_k$ are zero with probability one.

Having this approximate underlying distribution for U it is possible to define from it approximate distributions for other statistics based on U.

As a first example consider the angle θ between a pair of vectors U and U' independently distributed according to (i) and (ii) above. Due to the spherical symmetry of the distribution in r-space the conditional distribution of θ given U' does not depend on the particular U' so that the distribution of θ is the distribution of the angle between U and any fixed direction e.g.,

$$y_1 = 1, y_2 = y_3 = \cdots = y_k = 0.$$

Thus $\cos^2 \theta$ is distributed as

$$y_1^2/(y_1^2 + y_2^2 + \cdots + y_r^2)$$

i.e. $\cos^2 \theta$ has the β distribution $\beta_{1/2,(r-1)/2}$ defined by density

$$\frac{\Gamma\left(\frac{r}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{r-1}{2}\right)} x^{\frac{1}{2}-1}(1-x)^{(r-1)/2-1}.$$

This will be used as an approximation to the distribution of $\cos^2 \theta$ under the circumstances where $\mu\chi_r^2$ is used as an approximation to the distribution of Q.

Accepting these approximations it is natural to attempt to estimate μ and r. In particular, estimation of r plays a significant role in our non-exact test. The distribution theory leading to estimates of r will now be discussed. The vectors $V_1, V_2, \cdots, V_m$ may be described by the set of their lengths and the set of angles between pairs of them, and under the sphericalizing approximation these 2 sets of random variables are independent of one another. From each of these sets a statistic is defined which contains nearly all the information about r in the set and whose distribution may be approximated by a fast-converging limiting form as $r \to \infty$, namely $[(1/r) + (c/r^2)]\chi_h^2$ where c and h must be determined for each set. This leads to 2 independent estimates of r which may be combined into a single estimate.

Taking the set of lengths define Q_i as the squared length of U_i and consider $Q_i, \cdots, Q_m$ as m independent observations from $\mu\chi_r^2$ with μ and r unknown. The results of this paragraph are found in [7]. The joint density function of Q_1, $Q_2, \cdots, Q_m$ is

$$\left(2^{r/2}\Gamma\left(\frac{r}{2}\right)\mu^{r/2}\right)^{-m}\left(\prod_{i=1}^{m} Q_i\right)^{(r/2)-1} \exp\left(-\frac{1}{2\mu}\sum_{i=1}^{m} Q_i\right)$$

so that $\prod_{i=1}^m Q_i$ and $\sum_{i=1}^m Q_i$ are a pair of sufficient statistics for μ and r. It is now natural to look at

$$v = \frac{\prod_{i=1}^m Q_i}{(\sum_{i=1}^m Q_i)^m}$$

as a statistic not involving μ for the purpose of estimating r. From joint characteristic functions v and $\sum_{i=1}^{m} Q_i$ are seen to be independent. Thus

$$v \cdot \left(\sum_{i=1}^{m} Q_i\right)^m = \prod_{i=1}^{m} Q_i$$

where the 2 factors on the left are independent as are the m factors on the right. Since the distributions of $\sum_{i=1}^{m} Q_i$ and Q_i are known this equation makes it possible to immediately write down the moments of v about 0 or the cumulants and characteristic function of $\log v$. In this way we approach the limiting χ^2 distribution of $\log v$ as $r \to \infty$ and show that the power series expansion in terms of $(1/r)$ of the cumulants of the actual and asymptotic distributions agree up to the terms in $(1/r)^2$. This asymptotic distribution is stated in [7] to be remarkably good with agreement of the first 4 cumulants to within 5% when r is as small as 5.

Asymptotic expansions for the cumulants may be derived as follows. Define $t = -\log(m^m v)$, and K_s as meaning sth cumulant. Then for any s

$$K_s(\log v) + m^s K_s\left(\log \sum_{i=1}^{m} Q_i\right) = mK_s(\log Q_i),$$

or

$$K_s(\log v) + m^s K_s(\log \chi^2_{mr}) = mK_s(\log \chi^2_r).$$

From [8] asymptotic formulas for the cumulants of $\log \chi^2_n$ are given by

$$\begin{aligned} K_1(\log \chi^2_n) &= \log n - \frac{1}{n} - 2\sum_{j=1}^{\infty} \frac{(-4)^{j-1}B_j}{jn^{2j}} \\ &= \log n - \frac{1}{n} - \frac{1}{3n^2} + \frac{1}{15n^4} - \frac{16}{63n^6} + \cdots, \end{aligned}$$

and

$$\begin{aligned} K_s(\log \chi^2_n) &= (-)^s 2^s\left[\frac{(s-2)!}{2n^{s-1}} + \frac{(s-1)!}{2n^s} + \frac{2}{n^s}\sum_{j=1}^{\infty} \frac{(-4)^{j-1}B_j(2j+s-1)!}{(2j)!\,n^{2j-1}}\right] \\ &= (-)^s 2^s\left[\frac{(s-2)!}{2n^{s-1}} + \frac{(s-1)!}{2n^s} + \frac{s!}{6n^{s+1}} + \frac{0}{n^{s+2}} + \cdots\right] \quad \text{for} \quad s \geqq 2, \end{aligned}$$

where B_j are Bernoulli numbers. Thence

$$\begin{aligned} K_1(t) &= -m\log m - K_1(\log v) \\ &= -m\log m - mK_1(\log \chi^2_r) + mK_1(\log \chi^2_{mr}) \\ &= (m-1)\left[\frac{1}{r} + \frac{1+\frac{1}{m}}{3r^2} - \frac{2\left(1+\frac{1}{m}+\frac{1}{m^2}+\frac{1}{m^3}\right)}{15r^4} + \cdots\right], \end{aligned}$$

and for $s \geqq 2$

$$\begin{aligned} K_s(t) &= (-)^s K_s(\log v) = (-)^s[mK_s(\log \chi^2_r) - m^s K_s(\log \chi^2_{mr})] \\ &= 2^{s-1}(s-1)!(m-1)\left[\frac{1}{r^s} + \frac{s\left(1+\frac{1}{m}\right)}{3r^{s+1}} + \frac{0}{r^{s+2}} + \cdots\right] \end{aligned}$$

Since χ^2_{m-1} has cumulants $K_s = 2^{s-1}(s-1)!\,(m-1)$ it is seen that $t \sim (1/r)\chi^2_{m-1}$ with agreement in first terms of the expansions, and

$$t \sim \left(\frac{1}{r} + \frac{1 + \frac{1}{m}}{3r^2}\right)\chi^2_{m-1}$$

with agreement in the first 2 terms, for all cumulants.

Thus r may be estimated by $\hat{r}$ defined by

$$t = \left(\frac{1}{\hat{r}} + \frac{1 + \frac{1}{m}}{3\hat{r}^2}\right)(m-1)$$

and for r moderately large the distribution χ^2_{m-1} can be used to put confidence limits on r.

Consider next the set of $\frac{1}{2}m(m-1)$ angles among $U_1, U_2, \cdots, U_m$. Set $n = \frac{1}{2}m(m-1)$ and denote by S_i $(i = 1, 2, \cdots, n)$ the squared sines of these angles. Under the approximate model any S_i is considered distributed as $\beta_{(r-1)/2,1/2}$, but as a further consequence of spherical symmetry in r-space it may be noted that any set of angles containing no closed subset is a mutually independently distributed set, and in particular the angles are pairwise independent. Extending this approximation to complete independence the joint density of the S_i becomes

$$\left(\frac{\Gamma\left(\frac{r}{2}\right)}{\sqrt{\pi}\,\Gamma\left(\frac{r-1}{2}\right)}\right)^n \prod_{i=1}^n (S_i)^{(r-3)/2} \prod_{i=1}^n (1 - S_i)^{-\frac{1}{2}}$$

so that $\prod_{i=1}^n S_i$ or $\sum_{i=1}^n \log S_i$ appear as equivalent sufficient statistics for r, so contain approximately all the information about r in the directional properties.

This leads to a consideration of $-\log \beta_{(r-1)/2,1/2}$. The density of $1 - \beta_{(r-1)/2,1/2}$ is easily seen to be asymptotically as $r \to \infty$ the density of $1/r\ \chi_1^2$ and since $\beta_{(r-1)/2,1/2} \to 1$ in probability as $r \to \infty$ it follows that

$$\frac{-\log \beta_{(r-1)/2,1/2}}{1 - \beta_{(r-1)/2,1/2}} \to 1$$

in probability as $r \to \infty$ so that $-\log \beta_{(r-1)/2,1/2}$ is also asymptotically distributed as $1/r\ \chi_1^2$.

Direct asymptotic expansions for the cumulants of $-\log \beta_{(r-1)/2,1/2}$ show that, as with statistic t, this last asymptotic distribution can be modified to have agreement in the first 2 terms. For, since $\chi^2_{r-1} = \beta_{(r-1)/2,1/2} \cdot \chi^2_r$ with independence on the right (as may be seen by computing the characteristic functions of the logs of these random variables),

$$\begin{aligned} K_s\,(-\log \beta_{(r-1)/2,1/2}) &= (-)^s K_s\,(\log \beta_{(r-1)/2,1/2}) \\ &= (-)^s [K_s\,(\log \chi^2_{r-1}) - K_s\,(\log \chi^2_r)] \end{aligned}$$

for all s. Thence

$$K_1(-\log \beta_{(r-1)/2,1/2}) = \log r + \left[-\frac{1}{r} - \frac{1}{3r^2} + \frac{2}{15r^4} + \cdots\right]$$

$$- \log(r-1) - \left[-\frac{1}{r-1} - \frac{1}{3(r-1)^2} + \frac{2}{15(r-1)^4} + \cdots\right]$$

$$= \frac{1}{r} + \frac{3}{2r^2} + \frac{2}{r^3} + \frac{9}{4r^4} + \cdots,$$

and for $s \geqq 2$

$$K_s(-\log \beta_{(r-1)/2,1/2}) = 2^s\left[\frac{(s-2)!}{2(r-1)^{s-1}} + \frac{(s-1)!}{2(r-1)^s} + \frac{s!}{6(r-1)^s} + 0 + \cdots\right]$$

$$- 2\left[\frac{(s-2)!}{2r^{s-1}} + \frac{(s-1)!}{2r^s} + \frac{s!}{6r^{s-1}} + 0 + \cdots\right]$$

$$= 2^{s-1}(s-1)!\left[\frac{1}{r^s} + \frac{3s}{2r^{s+1}} + \frac{s(s+1)}{r^{s+2}} + \cdots\right]$$

so that

$$-\log \beta_{(r-1)/2,1/2} \sim \left(\frac{1}{r} + \frac{3}{2r^2}\right)\chi_1^2$$

with agreement to second terms in the expansions and therefore usable accuracy for quite small r.

Now we may regard

$$-\sum_{i=1}^{n} \log S_i \sim \left(\frac{1}{r} + \frac{3}{2r^2}\right)\chi_n^2$$

and obtain a new estimate of r. Since in approximation the angles were more than pairwise independent the first 2 moments of this last are asymptotically faithful to the approximate model. The remaining moments however will be distorted slightly on account of non-independence in a way which is difficult to investigate.

Finally an estimate of r can be obtained from $t - \sum_{i=1}^{n} \log S_i$ regarded as asymptotically $(1/r)\ \chi^2_{m-1+n}$ or an appropriate refinement for small r.

4. The non-exact significance test. The question is discussed here of what can be had in the way of a significance test based on $F = Q_0/1/m\sum_{i=1}^{m} Q_i$ considered as $F_{r,mr}$ under the null hypothesis where r is unknown but estimated from a statistic w considered distributed as $f(r)\chi_n^2$ independent of F with $f(r)$ equal to $1/r$ or an asymptotically equivalent refinement of $1/r$. The point estimate of r found from the equation $w = f(r)\cdot n$ will be denoted by $\hat{r}$ and the term $100p\%$ confidence point of r will indicate the value of r satisfying $w = f(r)\chi^2_{n(p)}$ where $\chi^2_{n(p)}$ denotes the $100p\%$ point of χ_n^2. Similar notation will be used for percentage points of other distributions.

A statistical test may be termed exact if the distribution of the test statistic under the null hypothesis does not depend on any unknown parameters. If r

were known the statistic F would have this property and the natural test would be to regard F as significant if $F > F_{r,mr(.95)}$. (Assume for this discussion a standard 5% nominal significance level.) Since r is unknown any test based on F must be non-exact and the natural non-exact test appears to be to regard F as significant if $F > F_{\hat{r},m\hat{r}(.95)}$. This test can also be formulated in terms of quantities α and $\hat{\alpha}$. Define α as the significance level of the observed F as a function of the true parameter r, i.e. α satisfies $F = F_{r,mr(1-\alpha)}$. Similarly $\hat{\alpha}$ as a function of the observed statistics F and w can be determined from $F = F_{\hat{r},m\hat{\alpha}(1-\hat{r})}$. The unattainable exact test is that F is significant if $\alpha < .05$; the non-exact test defined is that F is significant if $\hat{\alpha} < .05$.

The non-exact test still has a significance level (or size or probability of type I error) but this is now a function of r. Denoting this function by $\gamma(r)$ we have

$$\begin{aligned}\gamma(r) &= \Pr(\hat{\alpha} < .05)\\ &= \Pr(F > F_{\hat{r},m\hat{r}(.95)})\\ &= \operatorname*{ave}_{w}\{\Pr(F > F_{\hat{r}(w),m\hat{r}(w)(.95)} \mid w)\}\end{aligned}$$

where F is distributed as $F_{r,mr}$. The last version of $\gamma(r)$ indicates how $\gamma(r)$ can be calculated for given r i.e. by averaging a set of fairly well tabled probabilities over a χ^2 distribution. The major interest of this section is to determine the relation between $\gamma(r)$ and the nominal significance level .05.

The distributions of α and $\hat{\alpha}$ can be compared by fixing α and looking at the variability of the corresponding $\hat{\alpha}$. This amounts to conditioning the various random variables by fixing F to produce the desired α, but leaving w unconditioned. For any fixed α, if r is known, percentage points of w can be translated into percentage points of $\hat{r}$ and thence to percentage points of $\hat{\alpha}$. These are denoted $(\hat{\alpha} \mid \alpha)_{(p)}$. Alternatively, for fixed α, r unknown, but w observed, confidence points for r can be translated into confidence points for α and these will also indicate how much $\hat{\alpha}$ varies about α.

Short of actually calculating $\gamma(r)$ for various values of r, m, n and .05, two arguments will be advanced to show that it is near .05. The first argument is to use a table to back up the belief that the disturbance caused by going from α to $\hat{\alpha}$ is not very great relative to the (0, 1) range of α and is well balanced with regard to direction, so that the unconditional distribution of $\hat{\alpha}$ is not much different from the uniform (0, 1) distribution of α. Table 1 shows quartiles of $(\hat{\alpha} \mid \alpha)$ for $m = 10$, $n = 64$, $\alpha = .05$ and .10, and $r = 6$ and ∞. This table indicates that the

TABLE 1

r	α	$(\hat{\alpha}\mid\alpha)_{(.25)}$	$(\hat{\alpha}\mid\alpha)_{(.75)}$
6	.050	.043	.057
	.100	.091	.107
∞	.050	.030	.070
	.100	.072	.125

disturbance in α caused by using $\hat{\alpha}$ is well-balanced near the 5% level and is a slight shift towards 0 near 10%. The indication is that $\gamma(r)$ is very near .05.

The second argument involves computing the non-trivial limit $\gamma(\infty) = \lim_{r\to\infty} \gamma(r)$. Define

$$(F_{r,mr} - 1)^+ = 0 \quad \text{if} \quad F_{r,mr} - 1 \leqq 0$$
$$= F_{r,mr} - 1 \quad \text{otherwise.}$$

As

$$r \to \infty \ F_{r,mr} \sim N\left(1, \frac{2}{r}\left\{1 + \frac{1}{m}\right\}\right)$$

so that

$$(F_{r,mr} - 1)^+ \sim \left(N\left(0, \frac{2}{r}\left\{1 + \frac{1}{m}\right\}\right)\right)^+$$

or

$$[(F_{r,mr} - 1)^+]^2 \sim 0 \quad \text{or} \quad \frac{2}{r}\left(1 + \frac{1}{m}\right)\chi_1^2$$

each with probability $\frac{1}{2}$. Similarly if $(1/\hat{r}) = (1/rn)\chi_n^2$ is put in for $1/r$,

$$[(F_{\hat{r},m\hat{r}} - 1)^+]^2 \sim 0 \quad \text{or} \quad \frac{2}{r}\left(1 + \frac{1}{m}\right)\chi_1^2 = \frac{2}{r}\left(1 + \frac{1}{m}\right)\frac{1}{n}\chi_n^2 \cdot \chi_1^2$$

each with probability $\frac{1}{2}$ where χ_n^2 and χ_1^2 are independent. From this

$$\gamma(\infty) = \lim_{r\to\infty} \Pr\,(F_{r,mr} > F_{\hat{r},m\hat{r}(.95)})$$
$$= \lim_{r\to\infty} \Pr\,([(F_{r,mr} - 1)^+]^2 > [(F_{\hat{r},m\hat{r}} - 1)^+]^2_{(.95)})$$
$$= \tfrac{1}{2} \Pr\,(\chi_1^2 > \frac{1}{n}[\chi_n^2 \cdot \chi_1^2]_{(.90)})$$

Now

$$\text{ave}\,\{\chi_1^2\} = 1 \quad \text{and} \quad \text{var}\,\{\chi_1^2\} = 2,$$

and

$$\text{ave}\left\{\frac{1}{n}\chi_n^2 \cdot \chi_1^2\right\} = 1 \quad \text{and} \quad \text{var}\left\{\frac{1}{n}\chi_n^2 \cdot \chi_1^2\right\} = 2 + \frac{6}{n},$$

which indicates strongly that

$$\left[\frac{1}{n}\chi_n^2 \cdot \chi_1^2\right]_{(.90)} > \left[\chi_1^2\right]_{(.90)}$$

i.e. $\gamma(\infty) < .05$ so that asymptotically the test is conservative as r gets large. Since the foregoing table indicates smaller spread in $(\hat{\alpha} \mid \alpha)$ for finite r than for $r = \infty$ we might hope that $\gamma(r)$ is as near .05 as $\gamma(\infty)$ is.

In any particular case the spread in $(\hat{\alpha} \mid \alpha)$ can be examined by computing confidence points of α from confidence points of r.

One feature of this test which might be regarded as a practical drawback is the non-uniqueness of the vectors $U_1, U_2, \cdots, U_m$. These vectors resulted from a choice of an orthogonal set of m d.f. chosen arbitrarily in a space of $m = n_1 + n_2 - 2$ dimensions. The symmetry of the normal distribution over any choice of an orthogonal set of d.f. assures that the distribution theory of the test holds for any such set, but there is no assurance that the observed statistics are unchanged by different choices. In fact it can be easily seen that $\sum_{i=1}^{m} Q_i$ is invariant under all choices so that $F = Q_0/(1/m) \sum_{i=1}^{m} Q_i$ is also. Thus it is only in the estimation of r that variations occur, and, since we have heuristic evidence of having used almost all the information about r in our estimates $\hat{r}$ and since the $\hat{r}$ plays only a secondary role, the non-uniqueness of the significance test should be of minor importance.

5. The sensitivity of the test. A natural parameter measuring separation of the 2 populations is distance between their means in a metric defined as follows from their second order moments. Suppose that the metric inserted into affine k-space from a priori information and used heretofore is denoted by G_1, and suppose that the ellipsoid in affine space which appears as the unit sphere in G_1 is denoted by E_1. If, in an affine coordinate system for k-space a sample point is represented by $k \times 1$ vector u and the corresponding $k \times k$ matrix of variances and covariances is Λ, then an ellipsoid E_2 can be defined as $u'\Lambda^{-1}u = 1$. It is easily seen that the same ellipsoid is defined by the same prescription in any affine coordinate system so that given the distribution over affine space E_2 is uniquely defined. E_2 may now be used to define a Euclidean metric G_2 in affine space as that metric in which E_2 appears to be the unit sphere (so that the distribution is sphericalized). Suppose the G_2-distance between population means is τ i.e. ν_0 has G_2-length τ. Then τ, which is also the ratio of mean difference to standard deviation for that linear combination of original variables which maximizes this ratio, may be taken as the parameter measuring difference between population means, and we would like to know if our test is sensitive to large τ.

Now ave$\{V_0\} = \nu_0$ with G_2-length τ so ave$\{U_0\} = \xi = (1/n_1 + 1/n_2)^{-1/2}\nu_0$ with G_2-length $(1/n_1 + 1/n_2)^{-1/2}\tau = \tau_1$ say. Denote by $Q_0(\xi)$ the squared G_1-length of U_0, by $P_0(\xi)$ the squared G_2-length of U_0 and by $R(U_0)$ the squared ratio of the radius of E_1 to the radius of E_2 both radii in the direction of U_0. (This ratio of lengths in one direction is independent of the particular metric.) Then

$$Q_0(\xi) = R(U_0)P_0(\xi). \tag{5.1}$$

Assuming normality the distributions appear in G_2 as spherical unit normals so that $P_0(\xi)$ has the non-central χ^2 distribution $\chi_k^2(\tau_1)$ defined as the distribution of $(v_1 + \tau_1)^2 + v_2^2 + \cdots + v_k^2$ where $v_1, v_2, \cdots, v_k$ are NID(0, 1). Unfortunately $R(U_0)$ has a distribution depending on the direction of ξ as well as its length.

This may be contrasted with the statistic T^2 usable if $k \leqq m$, which is noncentrally distributed as

$$T^2 = \frac{m\chi_k^2(\tau_1)}{\chi_{m-k+1}^2}$$

with the numerator independent of the denominator [9] which distribution depends only on τ and on no other properties of ξ. The present situation has the undesirable feature that the G_1-metric may have been selected in such a way that the G_1-length of ξ is too small to cause a significant disturbance in Q_0 whereas τ_1 is large enough to cause a significant disturbance in $\chi_k^2(\tau_1)$. An extreme example of this would occur if the populations were not of full rank but lay in separate parallel hyperplanes but still very close in G_1. Here $\tau = \infty$ but $Q_0(\xi)$ could very well be little disturbed. As long as ξ is regarded as having non-random direction and G_1 cannot be chosen to coincide with G_2 there is danger of insensitivity to a large τ arising from this source. On the other hand if it is permissible to assume randomness for ξ then this danger can be controlled on the average, and further discussion proceeds along these lines.

The high-dimensional case is likely to arise in practice when little or nothing is known about the separating power of individual variables. If nothing is supposed known it may be reasonable to think of ξ as random with all directions intuitively equally likely. The only affine choice consistent with this intuitive notion is to make ξ uniformly distributed with respect to G_2-direction, so the first case considered will be where ξ has constant G_2-length τ_1 and is uniformly distributed with respect to G_2-direction independently of the within sample variation.

Under this assumption and normality $U_0(\xi) = \xi + U_0(0)$ where the 2 vectors on the right are independent each with directions distributed uniformly in G_2. Also, due to the G_2-spherical symmetry of the distribution of $U_0(0)$, the G_2-length of $U_0(0)$ is distributed independently of its G_2-direction. It follows that $U_0(\xi)$ has independently distributed G_2-length and G_2-direction, so that in the equation (5.1) the 2 terms on the right are independent. Also the distribution of $R(U_0)$ is independent of ξ. In our standard approximation of $Q_0(0)$ by $\mu\chi_r^2$ by fitting first 2 moments we have $\text{ave}\{Q_0(0)\} = \mu r$ and $\text{ave}\{Q_0^2(0)\} = \mu^2 r(r+2)$. Also $\text{ave}\{P_0(0)\} = \text{ave}\{\chi_k^2\} = k$ and $\text{ave}\{P_0^2(0)\} = k(k+2)$. Thus

$$\text{ave}\,\{R(U_0)\} = \frac{\text{ave}\,\{Q_0(0)\}}{\text{ave}\,\{P_0(0)\}} = \frac{\mu r}{k}$$

and

$$\text{ave}\,\{R^2(U_0)\} = \frac{\text{ave}\,\{Q_0^2(0)\}}{\text{ave}\,\{P_0^2(0)\}} = \frac{\mu^2 r(r+2)}{k(k+2)}$$

so that

$$\text{ave}\,\{Q_0(\xi)\} = \text{ave}\,\{R(U_0)\} \cdot \text{ave}\,\{\chi_k^2(\tau_1)\} = \mu r\left(1 + \frac{\tau_1^2}{k}\right),$$

$$\text{ave}\,\{Q_0^2(\xi)\} = \text{ave}\,\{R^2(U_0)\}\ \text{ave}\,\{\chi_k^4(\tau_1)\}$$
$$= \mu^2 r(r+2)\left(1 + 2\frac{\tau_1^2}{k} + \frac{\tau_1^4}{k(k+2)}\right),$$

and

$$\text{var}\,\{Q_0(\xi)\} = 2\mu^2 r\left(1 + 2\frac{\tau_1^2}{k} + \frac{k-r}{k+2}\frac{\tau_1^4}{k^2}\right).$$

The distribution of $Q_0(\xi)$ is clearly not non-central χ^2 since the variance of the latter does not involve a term in τ_1^4. For practical purposes it would be reasonable to fit a χ^2 shape to this distribution by fitting first 2 moments, i.e. $\lambda\chi_q^2$ where

$$\lambda = \mu\,\frac{1 + 2\dfrac{\tau_1^2}{k} + \dfrac{k-r}{k+2}\dfrac{\tau_1^4}{k^2}}{1 + \dfrac{\tau_1^2}{k}}$$

$$q = r\left(1 + \frac{\dfrac{r+2}{k+2}\dfrac{\tau_1^4}{k^2}}{1 + 2\dfrac{\tau_1^2}{k} + \dfrac{k-r}{k+2}\dfrac{\tau_1^4}{k^2}}\right)$$

Now it is possible to compute approximate "power functions" and "confidence limits" for τ by assuming F for $\tau > 0$ approximately distributed as $\lambda/\mu\ F_{q,mr}$ and by adopting the procedure used with significance testing of replacing r by $\hat{r}$. These "power functions" and "confidence limits" are actually estimates of the true power functions and confidence limits associated with the non-exact test just as $\hat{\alpha}$ was an estimate of α. The deviation of the estimated power from the true power may again be expected to be near zero and balanced about zero. Using confidence points of r confidence points for any particular value of the power function may be found and these will indicate the order of the disturbance caused by replacing r by $\hat{r}$.

For convenience a criterion different from the power function will be used to measure the sensitivity of the test, namely τ_c the value of τ which will produce on the average a barely significant test statistic. Regarding $(1/m)\sum_{i=1}^{m} Q_i$, the denominator of F, as $(\mu/m)\ \chi_{mr}^2$

$$\text{ave}\,\{F\} = \text{ave}\,\{Q_0(\xi)\}\ \cdot\ \text{ave}\left\{\frac{m}{\mu}\cdot\chi_{mr}^{-2}\right\}$$
$$= \mu r\left(1 + \frac{\tau_1^2}{k}\right)\cdot\frac{1}{\mu r}\frac{mr}{mr-2}$$
$$= \left(1 - \frac{2}{mr}\right)^{-1}\left(1 + \frac{\tau_1^2}{k}\right)$$

so that τ_c satisfies

$$1 + \left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{-1}\frac{\tau_c^2}{k} = \left(1 - \frac{2}{mr}\right)F_{r,mr(.95)}$$

or, since, for large r, $F_{r,mr} \sim N(1, (2/r)[1 + (1/m)])$, τ_c asymptotically satisfies

$$1 + \left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{-1} \frac{\tau_c^2}{k} = 1 + 1.65 \left(\frac{2}{r}\right)^{\frac{1}{2}} \left(1 + \frac{1}{m}\right)^{\frac{1}{2}}$$

or

$$\tau_c^2 = N \left(\frac{1}{n_1} + \frac{1}{n_2}\right) k r^{-\frac{1}{2}}$$

where $N = 1.65(2)^{\frac{1}{2}}(1 + (1/m))^{\frac{1}{2}} \doteq 2.3$. Note that for a given experiment $r^{-\frac{1}{2}}$ is the only factor in τ_c^2 which depends on G_1. This result is encouraging, for suppose we have a set of variables with equal but possibly small individual separation parameters ρ. If the within sample variation is independent from variable to variable then $\tau^2 = k\rho^2$. Thus if G_1 can be chosen such that

$$r \geqq \frac{N^2}{\rho^4} \left(\frac{1}{n_1} + \frac{1}{n_2}\right)^2$$

then separation would show on the average. This implies that regardless of how small ρ is we need only go on adding variables of separation ρ until r has been built up to correct size. Whether it is possible to continue indefinitely adding variables with small separations in a practical case is uncertain, but the example does show how small individual separations can produce something that will show.

If there is some feeling that ξ is not uniformly distributed relative to G_2-direction an alternative would be to suppose it uniform relative to a different metric G_3 with ellipsoid E_3, i.e. when E_3 appears as the unit sphere ξ appears of length σ and uniform with regard to direction independent of $Q_0(0)$. Then a priori knowledge of the separating powers of the variables could be supposed to consist of some information about E_3. Suppose the mean square G_1-length of ξ is $A^2\sigma^2$ where A^2 depends only on E_3, and suppose the mean square G_1-length of the centrally distributed U_0 is $B^2 = \mu r = \text{ave}\{Q_0(0)\}$. Then

$$\text{ave } \{Q_0(\xi)\} = B^2 + A^2\sigma = \mu r \left(1 + \frac{A^2}{B^2}\sigma^2\right)$$

so that σ_c^2 producing significance on the average is given, in the asymptotic case by

$$\sigma_c^2 = N \left(\frac{1}{n_1} + \frac{1}{n_2}\right) \frac{A^2}{B^2} r^{-\frac{1}{2}}$$

where now the choice of G_1 can influence both B/A and r.

We are now in a position to discuss theoretical issues concerning the original choice of metric G_1. These suggest that for most purposes the aim should be to make G_2 and G_1 coincide as nearly as possible except for a scale factor. The practical question of how well this can be accomplished is not discussed, nor is it crucial for the use of the method. There are 2 issues in the choice of G_1 : sensitivity of the test and safety of the assumptions.

If G_1 is related to G_2 by a scale factor, then the statistic Q is distributed as $\mu\chi_k^2$ i.e., $r = k$ and under normality the approximation to the distribution of Q by $\mu\chi_r^2$ is exact. This is the way in which choice of G_1 can be made to improve the assumptions. It is heuristically evident that a larger value of r results in less likelihood that approximations of this kind will go wrong.

Regarding sensitivity it can be seen that only when E_2 is G_1-spherical is there equal sensitivity to a separation of τ in all directions and so no danger of insensitivity to large τ. Also it has been seen that when the direction of ξ is assumed G_2-uniformly distributed τ_c^2 depends on $r^{-\frac{1}{2}}$ so again there is evidence that maximizing r to k gives greatest sensitivity. However, under the alternative randomness assumption of ξ uniform over ellipsoid E_3 the situation appears more complicated, for the factor in σ_c^2 to be minimized by choice of G_1 is $A^2/(B^2)\ r^{-\frac{1}{2}}$. This suggests that if something is known about the shape of E_3 as well as E_2 then E_1 should be chosen to give more weight to those directions in which E_3 is long relative to E_2 provided this does not too greatly depress r. It is felt that this last suggestion may be occasionally useful but the general rule will be to try to make $r = k$.

6. Asymptotic behavior. In the foregoing are many results asymptotically true as $r \rightarrow \infty$ with m fixed. Certainly these are a mathematical convenience. The question of whether indefinitely large r can be practically obtained remains open. Certainly if k can be made arbitrarily large and each of the k variables contains a part independent of the rest then in theory r can be made arbitrarily large because a metric can be chosen such that $r = k$. What is much more in doubt is whether or not variables could be chosen which would give r indefinitely increasing and τ also increasing at a rate such that the sensitivity of our method would continue to improve.

Whether it is practically attainable or just mathematically useful the following geometrical picture of the asymptotic case is illuminating. Consider throughout the approximate model of section 3 and its asymptotic behavior. As $r \rightarrow \infty$ the coefficient of variation of Q (i.e. $\mu\chi_r^2$) tends to 0, so that if we back away from the picture at the correct rate as r increases the vectors $U_1, \cdots, U_m$ will appear to all approach in probability the same constant length. Also since $1 - S_i \sim 1/(r)\,\chi_1^2$ each angle between vectors tends in probability to $\pi/2$ so they tend to an orthogonal set of m equal length vectors. Vector U_0 also becomes perpendicular to $U_1, \cdots, U_m$ but its length depends on τ. However if its length should differ from the common limiting lengths of the rest by a factor as great as $(1 + Nr^{-\frac{1}{2}})^{\frac{1}{2}}$ this is roughly what would be called significant, so that asymptotically a significant U_0 could be indistinguishably different from the rest.

An implication of this asymptotic picture is as follows. For small r it would be natural to compare $Q_0(\xi)$ from U_0 more closely with those Q_i from U_i making the smallest angles with U_0, because if U_0 and U_i are close then $R(U_0)$ and $R(U_i)$ are likely to be more nearly the same. T^2 accomplishes this in a neat manner which disappears when $k > m$, but the present method makes no attempt to do

it. The asymptotic picture says that in the limit there is no hope of making such a correction, for if U_0 is nearly at right angles with every U_i then the radii of E_1 and E_2 in the direction of U_0 bear no relation to the radii in the direction of the U_i, i.e. there are too many directions for U_0 to take to hope that it will be near enough to any U_i to make any difference.

7. Acknowledgments. Heartiest thanks are due to Professor John W. Tukey of Princeton University for his generous guidance in this research.

REFERENCES

[1] A. P. DEMPSTER, "The multivariate two sample problem in the degenerate case," unpublished PhD. thesis, Princeton University, 1956.

[2] H. HOTELLING, "The generalisation of 'Student's' ratio," *Ann. Math. Stat.*, Vol. 2 (1931), p. 360.

[3] R. A. FISHER, "The use of multiple measurements in taxonomic problems," *Annals of Eugenics*," Vol. 7 (1936), p. 179.

[4] E. J. G. PITMAN, "Significance tests which may be applied to samples from any population," J. Roy. Stat. Soc., Supplement, Vol. 4 (1937), p. 119-130, 225-232.

[5] B. L. WELCH, "On the z-test in randomized blocks and latin squares," *Biometrika*, Vol. 29 (1937), p. 21.

[6] G. E. P. BOX, "Some theorems on quadratic forms applied in the study of analysis of variance I," *Ann. Math. Stat.*, Vol. 25 (1954), p. 290.

[7] M. S. BARTLETT AND M. G. KENDALL, "The statistical analysis of variance heterogeneity and the logarithmic transformation," J. Roy. Stat. Soc., Supplement, Vol. 8 (1946), p. 128.

[8] J. WISHART, "The cumulants of the Z and of the logarithmic χ^2 and t distributions," *Biometrika*, Vol. 34 (1947), p. 170.

[9] R. C. BOSE AND S. N. ROY, "The distribution of Studentized D^2 statistic," *Sankhya*, Vol. 4 (1938), p. 337.

ON THE KOLMOGOROV AND SMIRNOV LIMIT THEOREMS FOR DISCONTINUOUS DISTRIBUTION FUNCTIONS

By Paul Schmid

Swiss Forest Research Institute and Federal Institute of Technology

1. Introduction. Let $X_1, X_2, \ldots, X_N$ be N independent random variables with the same distribution function $F(x)$. $S_N(x)$ is the empirical distribution function, i.e., $S_N(x) = k/N$ if exactly k of the N values X_i are less than or equal to x. It is of theoretical and practical interest to analyze the behavior of the statistics

$$\sup_{-\infty < x < \infty} |S_N(x) - F(x)| \cdot N^{\frac{1}{2}}$$

and

$$\sup_{-\infty < x < \infty} (S_N(x) - F(x)) \cdot N^{\frac{1}{2}}.$$

Kolmogorov [12] proved in a famous paper in 1933 that for $\lambda > 0$

$$\text{I} \qquad \lim_{N\to\infty} P[\sup_{-\infty < x < \infty} |S_N(x) - F(x)| \cdot N^{\frac{1}{2}} < \lambda] = \sum_{k=-\infty}^{+\infty} (-1)^k e^{-2\lambda^2 k^2}$$

if $F(x)$ is a continuous distribution function. Smirnov [21] obtained a similar result in 1939, when he showed that

$$\text{II} \qquad \lim_{N\to\infty} P[\sup_{-\infty < x < \infty} (S_N(x) - F(x)) \cdot N^{\frac{1}{2}} < \lambda] = 1 - e^{-2\lambda^2}$$

holds for continuous distribution functions $F(x)$.

Kolmogorov converts in his proof to a generalization of the Central Limit theorem, whereas Smirnov's theorem was a corollary to a more intricate theorem. But the two formulae can be proved by reciprocal methods. They have also been proved by Feller [11] and by Doob [10] and Donsker [9]. Feller made use of characteristic functions and Doob employed stochastic processes. Smirnov [22] found in 1944 the first terms of the asymptotic expansion for the probability in II and an exact formula for finite N. Chung [7] and Blackman [5], [6] were successful in finding the asymptotic expansion for the probability in I.

A somewhat more general form of the statistics, namely

$$\sup_{-\infty < x < \infty} |S_N(x) - F(x)| \, N^{\frac{1}{2}} \cdot \varphi(F(x)),$$

where $\varphi(y)$ is a positive definite weight function, was discussed by Anderson and Darling [1]. They found the limit distributions for some special weight functions,

Received October 11, 1957; revised June 20, 1958.

by means of stochastic processes. Similar results were obtained by Maniya [16] and Malmquist [15]. Rényi [19], in 1953, established the relations

$$\text{III} \quad \lim_{N\to\infty} P\left[\sup_{a \leq F(x)} \left|\frac{S_N(x) - F(x)}{F(x)}\right| \cdot N^{\frac{1}{2}} < \lambda\right] = \frac{4}{\pi}\sum_{k=0}^{\infty}(-1)^k \frac{\exp\left[-\dfrac{(2k+1)^2\pi^2}{8}\dfrac{1-a}{a\lambda^2}\right]}{2k+1}$$

and

$$\text{IV} \quad \lim_{N\to\infty} P\left[\sup_{a \leq F(x)} \left(\frac{S_N(x) - F(x)}{F(x)}\right) N^{\frac{1}{2}} < \lambda\right] = \sqrt{\frac{2}{\pi}}\int_0^{\lambda[a/(1-a)]^{\frac{1}{2}}} e^{-t^2/2}\, dt,$$

where $F(x)$ is a continuous distribution function, $a > 0$, $\lambda > 0$.

The statistics treated here are well suited to test if a sample comes from a population with the distribution function $F(x)$. These test functions have the great advantage in that their distributions are independent of the distribution $F(x)$ of the population. Massey [17], Birnbaum [2], and Malmquist [15] investigated the power of the statistics of Kolmogorov and Smirnov. The limit distributions of these statistics have been tabulated by Smirnov [23], and the distribution for finite N by Massey [18], Birnbaum and Tingey [3], [4]. Rényi tabulated his own limit distributions. Hence, today it is practicable to use these statistics.

In this paper Theorems I through IV are extended for the case of discontinuous distribution functions $F(x)$. The probabilities in question converge also in this case, but the limit distributions are no longer independent of $F(x)$. They depend on the values of $F(x)$ at the discontinuity points, but not on the form of the function between the points of discontinuity. Theorems 1 and 2 are proved by a generalization of the method of Kolmogorov. They can also be proved with the help of stochastic processes, as Doob did it for the case of continuous $F(x)$. We bypass representation of this method since it involves techniques similar to those of Anderson and Darling. The proofs of Theorems 3 and 4 follow in part the methods applied by Rényi, but also make use of the generalization by Kolmogorov of the Central Limit theorem. A part of these results has already been published [20].

I should like to thank W. Saxer for suggesting this topic.

2. Extension of the limit theorems of Kolmogorov and Smirnov. Let $F(x)$ be a distribution function continuous for $x \neq x_\nu$, where $F(x_\nu - 0) = f_{2\nu-1}$, $F(x_\nu) = f_{2\nu}$, for $\nu = 1, 2, \cdots, n$, and $f_{2n+1} = 1$. Denote the corresponding empirical distribution function by $S_N(x)$.

THEOREM 1. *If* $\lambda > 0$, *then*

$$(1) \quad \lim_{N\to\infty} P\left[\sup_{-\infty < x < \infty} |S_N(x) - F(x)| < \lambda N^{-\frac{1}{2}}\right] = \Phi(\lambda),$$

$$(2) \quad \Phi(\lambda) = \sum_{k=-\infty}^{+\infty}(-1)^k e^{-2\lambda^2 k^2}\, c \int \cdots \int_{G_k} \exp\left[-\frac{1}{2}\sum_{i,j=1}^{2n} \Lambda_{ij}\, x_i\, x_j\right] dx_1 \cdots dx_{2n},$$

where

$$\Lambda_{jj} = \frac{f_{j+1} - f_{j-1}}{(f_{j+1} - f_j)(f_j - f_{j-1})}, \qquad \Lambda_{jj-1} = \Lambda_{j-1j} = \frac{-1}{f_j - f_{j-1}},$$

$$\Lambda_{ij} = 0, \quad \text{for} \quad i < j - 1 \quad \text{or} \quad i > j + 1,$$

$$c = (2\pi)^{-n} \prod_{j=1}^{2n+1} (f_j - f_{j-1})^{-\frac{1}{2}}$$

and

$$G_k = \bigcup_{p_1, p_2, \cdots, p_n = -\infty}^{+\infty} \{-\lambda < x_{2\nu-1} + 2\lambda(p_\nu + kf_{2\nu-1}) < \lambda,$$

$$-\lambda < x_{2\nu} + 2\lambda(p_\nu + kf_{2\nu}) < \lambda, \qquad \nu = 1, \cdots, n\}.$$

THEOREM 2. *If* $\lambda > 0$, *then*

$$\lim_{N\to\infty} P[\sup_{-\infty<x<\infty} (S_N(x) - F(x)) < \lambda N^{-\frac{1}{2}}] = \Phi^+(\lambda), \tag{3}$$

$$\lim_{N\to\infty} P[\sup_{-\infty<x<\infty} (F(x) - S_N(x)) < \lambda N^{-\frac{1}{2}}] = \Phi^+(\lambda), \tag{4}$$

$$\Phi^+(\lambda) = \sum_{k=0}^{1} (-1)^k e^{-2\lambda^2 k^2} c \int \cdots \int_{G_k^+} \exp\left[-\frac{1}{2} \sum_{i,j=1}^{2n} \Lambda_{ij}\, x_i\, x_j\right] dx_1 \cdots dx_{2n}, \tag{5}$$

where

$$G_k^+ = \bigcup_{p_1, \cdots, p_n = 0}^{1} \{-\infty < (-1)^{p_\nu}(x_{2\nu-1} + 2\lambda k \cdot f_{2\nu-1}) + 2\lambda p_\nu < \lambda,$$

$$-\infty < (-1)^{p_\nu}(x_{2\nu} + 2\lambda k f_{2\nu}) + 2\lambda p_\nu < \lambda, \qquad \nu = 1, \cdots, n\}.$$

For $\lambda \leqq 0$ all limits are 0. The convergence is uniform in λ in all cases.

If the number of jumps of $F(x)$ is countably infinite, a further limit process has to be made in which at first only the highest jumps of $F(x)$ are taken into account. The two limit processes can be interchanged, because $\Phi(\lambda)$ and $\Phi^+(\lambda)$ are continuous functions of the values of $F(x)$ at the points of discontinuity. Hence further difficulties do not arise in this, the most general case. We will prove Theorem 1 for the case of a distribution function for which the inequalities

$$f_{2\nu+1} > f_{2\nu}, \qquad \nu = 0, 1, \cdots, n,$$

are valid. The results must then hold for any distribution function with n jumps, because both sides of (1) depend continuously on the f's.

If the random variable X has the distribution function $F(x)$, then $Y = F(X)$ is also a random variable, the distribution of which has to fulfill

$$P[F(X) \leqq 0] = 0, \qquad P[F(X) \geqq 1] = 0, \qquad P[f_{2\nu-1} \leqq F(X) < f_{2\nu}] = 0$$

and, for $f_{2\nu} \leqq y \leqq f_{2\nu+1}$,

$$P[F(X) \leqq y] = P[X \leqq F^{-1}(y)] = F(F^{-1}(y)) = y.$$

Furthermore, since

$$P[F(X) = f_{2\nu}] = P[X = x_\nu] = f_{2\nu} - f_{2\nu-1},$$

Y will have the distribution function

$$(6) \qquad F^0(y) = \begin{cases} 0, & \text{for } y \leqq 0, \\ y, & \text{for } f_{2\nu} \leqq y \leqq f_{2\nu+1}, \quad \nu = 0, 1, \cdots, n, \\ f_{2\nu-1}, & \text{for } f_{2\nu-1} \leqq y < f_{2\nu}, \quad \nu = 1, 2, \cdots, n, \\ 1, & \text{for } y \geqq 1. \end{cases}$$

Let $S_N^0(y)$ be the empirical distribution function corresponding to $F^0(y)$. Then we have, for $f_{2\nu} \leqq F(x) \leqq f_{2\nu+1}$,

$$\begin{aligned} S_N^0(F(x)) &= \frac{1}{N} \text{ (Number of } F(X_i), F(X_i) \leqq F(x)) \\ &= \frac{1}{N} \text{ (Number of } X_i, X_i \leqq x) = S_N(x) \end{aligned}$$

and $F^0(F(x)) = F(x)$. Hence

$$\sup_{-\infty<x<\infty} |S_N(x) - F(x)| = \sup_{-\infty<x<\infty} |S_N^0(x) - F^0(x)|,$$

because the other values of $F(x)$ cannot be attained. If we denote by I the union of the closed intervals $[f_{2\nu}, f_{2\nu+1}]$, $\nu = 0, 1, \cdots, n$, we obtain

$$(7) \quad P[\sup_{-\infty<x<\infty} |S_N(x) - F(x)| < \lambda N^{-\frac{1}{2}}] = P[\sup_{x \varepsilon I} |S_N^0(x) - x| < \lambda N^{-\frac{1}{2}}].$$

Denote by M_N the set of integers j such that $j/N \varepsilon I$,

$$(8) \quad M_N = \{k_0 = 0, 1, \cdots, k_1; k_2, k_2 + 1, \cdots, k_3; \\ \cdots; k_{2n}, k_{2n} + 1, \cdots, k_{2n+1} = N\}.$$

The k_i are defined such that $k_i/N \to f_i$, as $N \to \infty$. We wish to analyze the behavior of

$$(9) \qquad P\left[\max_{j \varepsilon M_N} \left| S_N^0\left(\frac{j}{N}\right) - \frac{j}{N} \right| < \lambda N^{-\frac{1}{2}}\right]$$

when $N \to \infty$.

The event $\mathcal{E}_{ik}$, $k \varepsilon M_N$, happens if simultaneously all inequalities

$$\left| S_N^0\left(\frac{j}{N}\right) - \frac{j}{N} \right| < \lambda N^{-\frac{1}{2}}, \qquad \text{for } j \leqq k, \ j \varepsilon M_N,$$

and the equality

$$S_N^0\left(\frac{k}{N}\right) - \frac{k}{N} = \frac{i}{N}$$

are fulfilled. P_{ik} is the probability of $\mathcal{E}_{ik}$. P_{0N} is equal to the probability in (9).

We can calculate the P_{ik} recursively by means of the initial conditions $P_{00} = 1$, $P_{i0} = 0$ for $i \neq 0$, and the equations

$$
\begin{aligned}
P_{i\,k+1} &= \sum_j P_{jk}\, P[\mathcal{E}_{i\,k+1} \mid \mathcal{E}_{jk}] \\
&= \sum_{|j|<\lambda N^{\frac{1}{2}}} P_{jk}\, P\left[S_N^0\left(\frac{k+1}{N}\right) - S_N^0\left(\frac{k}{N}\right)\right. \\
&\qquad \left. = \frac{i-j+1}{N} \,\middle|\, S_N^0\left(\frac{k}{N}\right) = \frac{k+j}{N}\right],
\end{aligned} \tag{10}
$$

for $k \, \varepsilon \, M_N$, $k+1 \, \varepsilon \, M_N$, and

$$
\begin{aligned}
P_{i\,k_{2\nu}} &= \sum_j P_{j\,k_{2\nu-1}}\, P[\mathcal{E}_{i\,k_{2\nu}} \mid \mathcal{E}_{i\,k_{2\nu-1}}] \\
&= \sum_{|j|<\lambda N^{\frac{1}{2}}} P_{j\,k_{2\nu-1}}\, P\left[S_N^0\left(\frac{k_{2\nu}}{N}\right) - S_N^0\left(\frac{k_{2\nu-1}}{N}\right)\right. \\
&\qquad \left. = \frac{i-j+k_{2\nu}-k_{2\nu-1}}{N} \,\middle|\, S_N^0\left(\frac{k_{2\nu-1}}{N}\right) = \frac{j+k_{2\nu-1}}{N}\right],
\end{aligned} \tag{11}
$$

for $\nu = 1, \cdots, n$.

The occurring conditional probabilities give

$$
\begin{aligned}
&P\left[S_N^0\left(\frac{k+1}{N}\right) - S_N^0\left(\frac{k}{N}\right) = \frac{i-j+1}{N} \,\middle|\, S_N^0\left(\frac{k}{N}\right) = \frac{k+j}{N}\right] \\
&\qquad = \binom{N-k-j}{i-j+1}\left(\frac{1}{N-k}\right)^{i-j+1}\left(\frac{N-k-1}{N-k}\right)^{N-k-i-1},
\end{aligned} \tag{12}
$$

for $k \, \varepsilon \, M_N$, $k+1 \, \varepsilon \, M_N$, and

$$
\begin{aligned}
&P\left[S_N^0\left(\frac{k_{2\nu}}{N}\right) - S_N^0\left(\frac{k_{2\nu-1}}{N}\right) = \frac{i-j+k_{2\nu}-k_{2\nu-1}}{N} \,\middle|\, S_N^0\left(\frac{k_{2\nu-1}}{N}\right) = \frac{k_{2\nu-1}+j}{N}\right] \\
&\qquad = \binom{N-k_{2\nu-1}-j}{i-j+k_{2\nu}-k_{2\nu-1}}\left(\frac{k_{2\nu}-k_{2\nu-1}}{N-k_{2\nu-1}}\right)^{i-j+k_{2\nu}-k_{2\nu-1}}\left(\frac{N-k_{2\nu}}{N-k_{2\nu-1}}\right)^{N-k_{2\nu}-i},
\end{aligned} \tag{13}
$$

for $\nu = 1, \cdots, n$, according to the laws of the binomial distribution.

The recursion formulae can be simplified if we introduce the new terms

$$
Q_{ik} = \frac{N^N (N-k-i)!}{N!(N-k)^{N-k-i} e^k} P_{ik}. \tag{14}
$$

Now we have

$Q_{00} = 1$; $\quad Q_{i0} = 0$, for $i \neq 0$; $\quad Q_{ik} = 0$, for $|i| \geqq \lambda N^{\frac{1}{2}}$;

$$
Q_{i\,k+1} = \sum_{|j|<\lambda N^{\frac{1}{2}}} Q_{jk} \frac{1}{(i-j+1)!} e^{-1},
$$

for $k \, \varepsilon \, M_N$ and $k+1 \, \varepsilon \, M_N$; $|i| < \lambda N^{\frac{1}{2}}$,

$$
Q_{i\,k_{2\nu}} = \sum_{|j|<\lambda N^{\frac{1}{2}}} Q_{j\,k_{2\nu-1}} \frac{(k_{2\nu}-k_{2\nu-1})^{i-k+k_{2\nu}-k_{2\nu-1}}}{(i-j+k_{2\nu}-k_{2\nu-1})!\, e^{k_{2\nu}-k_{2\nu-1}}}, \tag{15}
$$

for $\nu = 1, \cdots, n$, and for the probability (9) we obtain

$$P_{0N} = \frac{N!e^N}{N^N} Q_{0N}. \tag{16}$$

For finite N, P_{0N} is evaluable, but as $N \to \infty$ the number of necessary recursion steps tends to infinity.

Let Y_j, $j \in M_N$, be independent random variables with the distributions

$$P\left[Y_j = \frac{i-1}{\lambda N^{\frac{1}{2}}}\right] = \frac{1}{i!e}, \qquad i = 0, 1, 2, \cdots; j \neq k_{2\nu}, \tag{17}$$

$$P\left[Y_{k_{2\nu}} = \frac{i - k_{2\nu} + k_{2\nu-1}}{\lambda N^{\frac{1}{2}}}\right] = \frac{(k_{2\nu} - k_{2\nu-1})^i}{i!e^{k_{2\nu}-k_{2\nu-1}}}, \qquad i = 0, 1, 2, \cdots. \tag{18}$$

Then

$$E(Y_j) = 0,$$

$$E(Y_j^2) = \frac{1}{\lambda^2 N}, \quad j \neq k_{2\nu}; \qquad E(Y_{k_{2\nu}}^2) = \frac{k_{2\nu} - k_{2\nu-1}}{\lambda^2 N},$$

$$E(|Y_j|^3) = \left(1 + \frac{2}{e}\right)\frac{1}{\lambda^3 N^{\frac{3}{2}}}, \quad j \neq k_{2\nu}; \qquad E(|Y_{k_{2\nu}}|^3) \sim \sqrt{\frac{8}{\pi}}\, \frac{(k_{2\nu} - k_{2\nu-1})^{\frac{3}{2}}}{\lambda^3 N^{\frac{3}{2}}}.$$

The event $\mathfrak{D}_{ik}$, $k \in M_N$, take place if the inequalities

$$\left|\sum_{l \leq j} Y_l\right| < 1$$

for all $j \leq k$ and the equality

$$\sum_{l \leq k} Y_l = \frac{i}{\lambda N^{\frac{1}{2}}}$$

are simultaneously fulfilled. The probability of $\mathfrak{D}_{ik}$ is R_{ik}, $R_{00} = 1$, $R_{i0} = 0$ for $i \neq 0$.

We can easily verify that the recursion formulae for the R_{ik} are the same as for the Q_{ik}. Therefore,

$$R_{ik} = Q_{ik}, \tag{19}$$

for all i and k. For the probability in (9) we obtain

$$P_{0N} = \frac{N!e^N}{N^N} R_{0N}. \tag{20}$$

We can evaluate R_{0N} in $2n + 1$ recursion steps from

$$R_{00} = 1, \qquad R_{i0} = 0, \qquad \text{for } i \neq 0,$$

$$\begin{aligned} R_{ik_{2\nu+1}} &= \sum_{|j| < \lambda N^{\frac{1}{2}}} R_{jk_{2\nu}} P[\mathfrak{D}_{ik_{2\nu+1}} \mid \mathfrak{D}_{jk_{2\nu}}], & \nu = 0, \cdots, n, \\ R_{ik_{2\nu}} &= \sum_{|j| < \lambda N^{\frac{1}{2}}} R_{jk_{2\nu-1}} \frac{(k_{2\nu} - k_{2\nu-1})^{i-j+k_{2\nu}-k_{2\nu-1}}}{(i - j + k_{2\nu} - k_{2\nu-1})!\, e^{k_{2\nu}-k_{2\nu-1}}}, & \nu = 1, \cdots, n. \end{aligned} \tag{21}$$

These conditional probabilities can be written in the form

$$P[\mathfrak{D}_{ik_{2\nu+1}} \mid \mathfrak{D}_{jk_{2\nu}}] = P\left[-1 - \frac{j}{\lambda N^{\frac{1}{2}}} < \sum_{r=k_{2\nu}+1}^{l} Y_r < 1 - \frac{j}{\lambda N^{\frac{1}{2}}},\right. \tag{22}$$
$$\left. l = k_{2\nu} + 1, \cdots, k_{2\nu+1};\ \sum_{r=k_{2\nu}+1}^{k_{2\nu+1}} Y_r = \frac{i-j}{\lambda N^{\frac{1}{2}}}\right],$$

and their limits for $N \to \infty$ can be obtained by the following lemma of Kolmogorov.

LEMMA, [12]. *Let $Y_{M1}, \cdots, Y_{Mm_M}$ be, for each M, independent random variables, whose values are multiples of $\epsilon = \epsilon(M)$, with*

$$E(Y_{Mj}) = 0, \qquad E(Y_{Mj}^2) = 2b_{Mj}, \qquad E(|Y_{Mj}|^3) = d_{Mj}.$$

Let a and b be two numbers such that $a < 0$ and $b > 0$. Assume the existence of positive numbers $A, \cdots, E$, such that, for all M, the inequalities (i) *through* (iv) *are fulfilled*:

(i) $A < \sum_{j=1}^{m_M} b_{Mj} < B$,

(ii) $\dfrac{d_{Mj}}{b_{Mj}} < C\epsilon$, for all j,

(iii) $P[Y_{Mj} = l_{Mj}\epsilon] > D$ and $P[Y_{Mj} = (l_{Mj} + 1)\epsilon] > D$ for all j and suitably chosen l_{Mj},

(iv) $a + E < i_M\epsilon < b - E$.

Then

$$P\left[a < \sum_{k=1}^{j} Y_{Mk} < b, j = 1, 2, \cdots, m_M;\ \sum_{k=1}^{m_M} Y_{Mk} = i_M\epsilon\right]$$
$$= \epsilon\left(v\left(0, 0, i_M\epsilon, 2\sum_{k=1}^{m_M} b_{Mk}\right) + \Delta\right),$$

where $u(\sigma, \tau, s, t)$ is Green's function for the heat equation

$$\frac{\partial f}{\partial t} = \frac{\partial^2 f}{\partial s^2}$$

in the region G,

$$G = \{a < s < b, t > 0\}.$$

If $\epsilon(M) \to 0$, then $\Delta \to 0$.

This lemma can be applied to the random variables $Y_{k_{2\nu}+1}, Y_{k_{2\nu}+2}, \cdots, Y_{k_{2\nu+1}}$. It should be noticed that the variables $Y_{k_{2\nu}}$ do not fulfill condition (ii) and hence must be treated independently. Our recursion formulae are now

$$R_{00} = 1, \qquad R_{i0} = 0, \qquad i \neq 0,$$

$$R_{ik_{2\nu+1}} = \sum_{|j|<\lambda N^{\frac{1}{2}}} R_{jk_{2\nu}} \frac{1}{\lambda N^{\frac{1}{2}}}\left(u_j\left(0, 0; \frac{i-j}{\lambda N^{\frac{1}{2}}}, \frac{k_{2\nu+1} - k_{2\nu}}{2\lambda^2 N}\right) + \Delta\right), \qquad \nu = 0, \cdots, n, \tag{23}$$

$$R_{ik_{2\nu}} = \sum_{|j|<\lambda N^{\frac{1}{2}}} R_{jk_{2\nu-1}} \frac{(k_{2\nu} - k_{2\nu-1})^{i-j+k_{2\nu}-k_{2\nu-1}}}{(i - j + k_{2\nu} - k_{2\nu-1})!\,e^{k_{2\nu}-k_{2\nu-1}}}, \qquad \nu = 1, \cdots, n,$$

where $u_j(\sigma, \tau; s, t)$ is Green's function for the heat equation in the region G_j,

$$G_j = \left\{-1 - \frac{j}{\lambda N^{\frac{1}{2}}} < s < 1 - \frac{j}{\lambda N^{\frac{1}{2}}}, t > 0\right\},$$

or

$$\tag{24} \begin{aligned} u_j(\sigma, \tau; s, t) &= \frac{1}{2\sqrt{\pi(t-\tau)}} \sum_{l=-\infty}^{+\infty} (-1)^l \\ &\quad \cdot \exp\left[-\frac{\left(s + \frac{j}{\lambda N^{\frac{1}{2}}} - (-1)^l\left(\sigma + \frac{j}{\lambda N^{\frac{1}{2}}}\right) - 2l\right)^2}{4(t-\tau)}\right]. \end{aligned}$$

If N tends to infinity, the Δ's disappear and the sums over j go over into integrals, with the exception of the sum in the first step which consists of only one summand,

$$R_{jk_1} = \frac{1}{\lambda N^{\frac{1}{2}}}\left(u\left(0, 0: \frac{j}{\lambda N^{\frac{1}{2}}}, \frac{k_1}{2\lambda^2 N}\right) + \Delta\right).$$

With this exception all sums tend to finite positive limits. The factor in (20), multiplied by $N^{-\frac{1}{2}}$ also tends to a finite limit, namely

$$N^{-\frac{1}{2}} \frac{N!\,e^N}{N^N} \sim \sqrt{2\pi}.$$

For $r \cdot N^{-\frac{1}{2}} \to x$, we obtain

$$\frac{N^{\frac{1}{2}}(k_{2\nu} - k_{2\nu-1})^{r+k_{2\nu}-k_{2\nu-1}}}{(r + k_{2\nu} - k_{2\nu-1})!\,e^{k_{2\nu}-k_{2\nu-1}}} \sim \frac{1}{\sqrt{2\pi(f_{2\nu} - f_{2\nu-1})}} \exp\left[-\frac{1}{2}\frac{x^2}{f_{2\nu} - f_{2\nu-1}}\right].$$

Finally we have

$$\tag{25} \begin{aligned} &\lim_{N\to\infty} P\left[\max_{j\in M_N}\left|S_N^0\left(\frac{j}{N}\right) - \frac{j}{N}\right| < \lambda N^{-\frac{1}{2}}\right] \\ &\quad = \sum_{j_0, j_1, \cdots, j_n=-\infty}^{+\infty} (-1)^{\sum_{i=0}^n j_i}(2\pi)^{-n} \prod_{j=1}^{2n+1} (f_j - f_{j-1})^{-\frac{1}{2}} \\ &\quad \cdot \int_{-\lambda<x_j<\lambda} \cdots \int \exp\left[-\frac{1}{2}\sum_{\nu=1}^{n} \frac{(x_{2\nu} - x_{2\nu-1})^2}{f_{2\nu} - f_{2\nu-1}}\right. \\ &\qquad\qquad \left. - \frac{1}{2}\sum_{\nu=0}^{n} \frac{(x_{2\nu+1} - (-1)^{j_\nu}x_{2\nu} - 2\lambda j_\nu)^2}{f_{2\nu+1} - f_{2\nu}}\right] dx_1 \cdots dx_{2n}, \end{aligned}$$

where x_0 and x_{2n+1} should be replaced by 0. This expression is $\Phi(\lambda)$.

Let us now prove that for those values of λ and sequences of N for which $\lambda N^{\frac{1}{2}}$ are integers,

$$\tag{26} \lim_{N\to\infty} P[\sup_{x\in I} |S_N^0(x) - x| < \lambda N^{-\frac{1}{2}}] = \lim_{N\to\infty} P\left[\max_{j\in M_N}\left|S_N^0\left(\frac{j}{N}\right) - \frac{j}{N}\right| < \lambda N^{-\frac{1}{2}}\right]$$

must be true.

To each $x \varepsilon I$ there exists a $j \varepsilon M_N$ such that either $x = j/N$ or $x = j/N + \epsilon$ with $0 < \epsilon < 1/N$. Set $S_N^0(x) = i/N$. From

$$S_N^0(x) - x = \frac{i-j}{N} - \epsilon \geqq \frac{\lambda N^{\frac{1}{2}}}{N}$$

follows, for $\epsilon > 0$,

$$S_N^0\left(\frac{j+1}{N}\right) - \frac{j+1}{N} \geqq S_N^0(x) - \frac{j+1}{N} = \frac{i-j-1}{N} \geqq \frac{\lambda N^{\frac{1}{2}}}{N},$$

because the value to the right is a multiple of $1/N$. From

$$S_N^0(x) - x = \frac{i-j}{N} - \epsilon \leqq -\frac{\lambda N^{\frac{1}{2}}}{N}$$

follows analogously

$$S_N^0\left(\frac{j}{N}\right) - \frac{j}{N} \leqq S_N^0(x) - \frac{j}{N} = \frac{i-j}{N} \leqq -\frac{\lambda N^{\frac{1}{2}}}{N}.$$

The second probability in (26) cannot be smaller than the first one and the limit of the second probability depends continuously on the endpoints of the intervals of I. Therefore the two limits have to be equal. The convergence must be uniform in λ, since $\Phi(\lambda)$ is a bounded and continuous function. Hence

$$P\,[\sup_{x \varepsilon I} |\, S_N^0(x) - x \,| < \lambda N^{-\frac{1}{2}}]$$

tends to $\Phi(\lambda)$ for all λ and all sequences of N. In view of (7), this proves Theorem 1.

Theorem 2 can be proved in a similar manner. We now disregard the absolute value signs in the definition of $\mathcal{E}_{ik}$ and $\mathfrak{D}_{ik}$. The summations in (10), (11), (15), (21) and (23) go from $-\infty$ to $\lambda N^{\frac{1}{2}}$ and the lower boundaries for the partial sums in (22) are omitted. Green's function for the heat equation in the region G_j^+,

$$G_j^+ = \left\{ s < 1 - \frac{j}{\lambda N^{\frac{1}{2}}},\, t > 0 \right\},$$

is now

$$u_j^+(\sigma, \tau; s, t) = \frac{1}{2\sqrt{\pi(t-\tau)}} \sum_{l=0}^{1} (-1)^l \cdot \exp\left[\frac{-\left(s + \frac{j}{\lambda N^{\frac{1}{2}}} - (-1)^l\left(\sigma + \frac{j}{\lambda N^{\frac{1}{2}}}\right) - 2l\right)^2}{4(t-\tau)} \right].$$

Hence

$$\lim_{N\to\infty} P\left[\max_{j \varepsilon M_N} \left(S_N^0\left(\frac{j}{N}\right) - \frac{j}{N} \right) < \lambda N^{-\frac{1}{2}} \right] = (2\pi)^{-n} \prod_{l=1}^{2n+1} (f_l - f_{l-1})^{-\frac{1}{2}} \sum_{j_0, j_1, \cdots, j_n = 0}^{1} (-1)^{\sum_{i=0}^{n} j_i}$$

$$\cdot \int_{x_j<\lambda} \cdots \int \exp\left[-\frac{1}{2}\sum_{\nu=1}^{n}\frac{(x_{2\nu}-x_{2\nu-1})^2}{f_{2\nu}-f_{2\nu-1}}\right.$$

$$\left. -\frac{1}{2}\sum_{\nu=0}^{n}\frac{(x_{2\nu+1}-(-1)^{j_\nu}x_{2\nu}-2\lambda j_\nu)^2}{f_{2\nu+1}-f_{2\nu}}\right] dx_1 \cdots dx_{2n},$$

where again x_0 and x_{2n+1} are 0. This proves Theorem 2.

3. Extension of the limit theorems of Rényi. Let $F(x)$ be a continuous function for $x \neq x_\nu$, with $F(x_\nu - 0) = f_{2\nu-1}$ and $F(x_\nu) = f_{2\nu}$, for $\nu = 1, 2, \cdots, n$, and $f_{2n+1} = 1$. Let f_0 be a positive number such that $f_0 \leqq f_1$. If $f_0 > f_1$, then we get the same results except that only the $f_i \geqq f_0$ will appear. Denote the empirical distribution function by $S_N(x)$.

THEOREM 3. *If* $\lambda > 0$, *then*

$$\lim_{N\to\infty} P\left[\sup_{f_0 \leq F(x)}\left|\frac{S_N(x)-F(x)}{F(x)}\right| < \lambda N^{-\frac{1}{2}}\right] = \Psi(\lambda), \tag{27}$$

$$\Psi(\lambda) = \sum_{k=-\infty}^{+\infty}(-1)^k d \int \cdots \int_{H_k} \exp\left[-\frac{1}{2}\sum_{i,j=0}^{2n}\Lambda_{ij}x_i x_j\right] dx_0 \cdots dx_{2n}, \tag{28}$$

where

$$\Lambda_{jj} = \frac{(f_{j+1}-f_{j-1})f_j^2}{(f_{j+1}-f_j)(f_j-f_{j-1})}, \Lambda_{j-1\,j} = \Lambda_{jj-1} = \frac{-f_j f_{j-1}}{(f_j - f_{j-1})},$$

$$\Lambda_{ij} = 0, \qquad \textit{for } i < j-1 \quad \textit{or} \quad i > j+1,$$

$$d = (2\pi)^{-n-\frac{1}{2}}\prod_{j=0}^{2n}(f_{j+1}-f_j)^{-\frac{1}{2}}(f_{j+1}^{\frac{1}{2}}f_j^{\frac{1}{2}}),$$

and

$$H_k = \bigcup_{p_1,\cdots,p_n=-\infty}^{+\infty}\{-\lambda < (-1)^k x_0 + 2\lambda k < \lambda; -\lambda < (-1)^{p_\nu}x_{2\nu-1} + 2\lambda p_\nu < \lambda,$$

$$-\lambda < (-1)^{p_\nu}x_{2\nu} + 2\lambda p_\nu < \lambda, \nu = 1, \cdots, n\}.$$

THEOREM 4. *If* $\lambda > 0$, *then*

$$\lim_{N\to\infty} P\left[\sup_{f_0 \leq F(x)}\frac{S_N(x)-F(x)}{F(x)} < \lambda N^{-\frac{1}{2}}\right] = \Psi^+(\lambda), \tag{29}$$

$$\lim_{N\to\infty} P\left[\sup_{f_0 \leq F(x)}\frac{F(x)-S_N(x)}{F(x)} < \lambda N^{-\frac{1}{2}}\right] = \Psi^+(\lambda), \tag{30}$$

$$\Psi^+(\lambda) = \sum_{k=0}^{1}(-1)^k d \int_{H_k^+} \cdots \int \exp\left[-\frac{1}{2}\sum_{i,j}\Lambda_{ij}x_i x_j\right]_{dx_0\cdots dx_{2n}}, \tag{31}$$

where

$$H_k^+ = \bigcup_{p_1,\cdots,p_n=0}^{1}\{-\infty < (-1)^k x_0 + 2\lambda k < \lambda;$$

$$-\infty < (-1)^{p_\nu}x_{2\nu-1} + 2\lambda p_\nu < \lambda, -\infty < (-1)^{p_\nu}x_{2\nu} + 2\lambda p_\nu < \lambda, \nu = 1, \cdots, n\}.$$

The convergence is in both theorems uniform in λ and for $\lambda \leqq 0$ all limits are 0.

These theorems can also be extended for distribution functions with infinitely many points of discontinuity.

We introduce again the random variable $Y = F(X)$ with the distribution function $F^0(x)$ and the set I as the union of the intervals $[f_{2\nu}, f_{2\nu+1}]$, $\nu = 0, 1, \cdots, n$. For any $F(x) \varepsilon I$ we have

$$\frac{S_N(x) - F(x)}{F(x)} = \frac{S_N^0(F(x)) - F^0(F(x))}{F^0(F(x))},$$

and therefore

$$\sup_{F(x) \geqq f_0} \left| \frac{S_N(x) - F(x)}{F(x)} \right| = \sup_{x \varepsilon I} \left| \frac{S_N^0(x) - x}{x} \right|.$$

Let $R_N(x)$ be the empirical distribution function of a sample $Z_1, Z_2, \cdots, Z_N$ from a population with the distribution

$$P[Z \leqq x] = x, \qquad 0 \leqq x \leqq 1, \tag{32}$$

then

$$P\left[\sup_{x \varepsilon I} \left| \frac{S_N^0(x) - x}{x} \right| < \lambda N^{-\frac{1}{2}}\right] = P\left[\sup_{x \varepsilon I} \left| \frac{R_N(x) - x}{x} \right| < \lambda N^{-\frac{1}{2}}\right], \tag{33}$$

since the distributions of the two populations coincide for $x \varepsilon I$. Thus,

$$P\left[\sup_{F(x) \geqq f_0} \left| \frac{S_N(x) - F(x)}{F(x)} \right| < \lambda N^{-\frac{1}{2}}\right] = P\left[\sup_{x \varepsilon I} \left| \frac{R_N(x) - x}{x} \right| < \lambda N^{-\frac{1}{2}}\right]. \tag{34}$$

The set I_ϵ is defined as the union of the intervals $[f_{2\nu} - \epsilon, f_{2\nu+1} + \epsilon]$, $\nu = 0, 1, \cdots, n$, for $\epsilon > 0$. If $|R_N(x) - x| \leqq \epsilon$, then

$$\sup_{R_N(x) \varepsilon I} \left| \frac{R_N(x) - x}{x} \right| \leqq \sup_{x \varepsilon I_\epsilon} \left| \frac{R_N(x) - x}{x} \right|,$$

since $R_N(x) \varepsilon I$ implies that $x \varepsilon I_\epsilon$. We see that

$$\begin{aligned} &P\left[\sup_{x \varepsilon I_\epsilon} \left| \frac{R_N(x) - x}{x} \right| < \lambda N^{-\frac{1}{2}}\right] \\ &\qquad \leqq P[|R_N(x) - x| > \epsilon] + P\left[\sup_{R_N(x) \varepsilon I} \left| \frac{R_N(x) - x}{x} \right| < \lambda N^{-\frac{1}{2}}\right]. \end{aligned} \tag{35}$$

By a similar procedure we have

$$\begin{aligned} &P\left[\sup_{R_N(x) \varepsilon I_{2\epsilon}} \left| \frac{R_N(x) - x}{x} \right| < \lambda N^{-\frac{1}{2}}\right] \\ &\qquad \leqq P[|R_N(x) - x| > \epsilon] + P\left[\sup_{x \varepsilon I_\epsilon} \left| \frac{R_N(x) - x}{x} \right| < \lambda N^{-\frac{1}{2}}\right]. \end{aligned} \tag{36}$$

It is sufficient to prove

$$\lim_{N \to \infty} P\left[\sup_{R_N(x) \varepsilon I} \left| \frac{R_N(x) - x}{x} \right| < \lambda N^{-\frac{1}{2}}\right] = \Psi(\lambda) \tag{37}$$

since the probability

$$P[\,|R_N(x) - x| > \epsilon]$$

tends to 0 as $N \to \infty$. The function $\Psi(\lambda)$ is continuously dependent on the boundaries of I. Therefore, from (35), (36) and (37) we get

$$\lim_{N\to\infty} P\left[\sup_{x\epsilon I}\left|\frac{R_N(x) - x}{x}\right| < \lambda N^{-\frac{1}{2}}\right] = \Psi(\lambda) \tag{38}$$

and from (34) follows the statement of Theorem 3.

We arrange the numbers $Z_1, \cdots, Z_N$ of the sample according to their values and denote by Z_k^* the one for which there are exactly $k - 1$ smaller numbers in the sample. The probability of ties is 0 since (33) is a continuous distribution function. $R_N(x)$ is equal to k/N in $Z_k^* \leqq x \leqq Z_{k+1}^*$. In this interval

$$\sup_{Z_k^* \leqq x < Z_{k+1}^*}\left|\frac{R_N(x) - x}{x}\right| = \max\left\{\left|\frac{k/N}{Z_k^*} - 1\right|, \left|\frac{k/N}{Z_{k+1}^*} - 1\right|\right\}.$$

For $Z_{k+1}^* \geqq k/N$

$$\left|\frac{k/N}{Z_{k+1}^*} - 1\right| \leqq \left|\frac{(k+1)/N}{Z_{k+1}^*} - 1\right| + \frac{1}{f_0 N}$$

since $k/N \geqq f_0$ implies that $Z_{k+1}^* \geqq f_0$. For $Z_{k+1}^* < k/N$

$$\left|\frac{k/N}{Z_{k+1}^*} - 1\right| \leqq \left|\frac{(k+1)/N}{Z_{k+1}^*} - 1\right|.$$

Therefore,

$$\max_{k/N\epsilon I_{1/N}}\left|\frac{k/N}{Z_k^*} - 1\right| \leqq \sup_{R_N(x)\epsilon I_{1/N}}\left|\frac{R_N(x) - x}{x}\right| \leqq \max_{k/N\epsilon I_{2/N}}\left|\frac{k/N}{Z_k^*} - 1\right| + \frac{1}{f_0 N}$$

and (37) is equivalent to

$$\lim_{N\to\infty} P\left[\max_{k/N\epsilon I_{1/N}}\left|\frac{k/N}{Z_k^*} - 1\right| < \lambda N^{-\frac{1}{2}}\right] = \Psi(\lambda). \tag{39}$$

We can write this equation as

$$\lim_{N\to\infty} P\left[\max_{k/N\epsilon I_{1/N}}\left|\ln\left(\frac{k/N}{Z_k^*}\right)\right| < \lambda N^{-\frac{1}{2}}\right] = \Psi(\lambda) \tag{40}$$

or, since $\log n - \sum_{r=1}^{n} 1/r \to c$,

$$\lim_{N\to\infty} P\left[\max_{k/N\epsilon I_{1/N}}\left|\ln\frac{1}{Z_k^*} - \sum_{l=k}^{n}\frac{1}{l}\right| < \lambda N^{-\frac{1}{2}}\right] = \Psi(\lambda). \tag{41}$$

The random variables $\ln(1/Z_k^*)$ are not independent since they fulfill the inequalities

$$\ln\left(\frac{1}{Z_N^*}\right) < \ln\left(\frac{1}{Z_{N-1}^*}\right) < \cdots < \ln\left(\frac{1}{Z_1^*}\right).$$

However, they do form an additive Markov chain (cf [24]), i.e., their differences

$$\ln\left(\frac{1}{Z_{k-1}^*}\right) - \ln\left(\frac{1}{Z_k^*}\right)$$

are mutually independent. The variables

$$U_l = (N + 1 - l)\left(\ln\frac{1}{Z_{N+1-l}^*} - \ln\frac{1}{Z_{N+2-l}^*}\right), \qquad l = 1, \cdots, N,$$

have the distribution

$$P[U_l \leqq x] = 1 - e^{-x}, \qquad 0 \leqq x < \infty.$$

On the other hand we obtain

$$\ln\left(\frac{1}{Z_k^*}\right) = \sum_{l=1}^{N+1-k} \frac{U_l}{N + 1 - l}$$

and

$$\ln\left(\frac{1}{Z_k^*}\right) - \sum_{l=k}^{N}\frac{1}{l} = \sum_{l=1}^{N+1-k} \frac{U_l - 1}{N + 1 - l}.$$

Some moments of the variables

$$V_l = N^{\frac{1}{2}} \frac{U_l - 1}{N + 1 - l}$$

are

$$E(V_l) = 0, \qquad E(V_l^2) = \frac{N}{(N + 1 - l)^2}, \; E(|V_l|^3) = \left(\frac{12}{e} - 2\right)\frac{N^{\frac{3}{2}}}{(N + 1 - l)^3}.$$

Let the set of integers j, for which $(N + 1 - j)/N \,\varepsilon\, I_{1/N}$, be

$$\{j_0 = 0, 1, \cdots, j_1; j_2, j_2 + 1, \cdots, j_3; \cdots; j_{2n}, j_{2n} + 1, \cdots, j_{2n} + 1\}.$$

The j_i are defined such that $j_i/N \to 1 - f_{2n+1-i}$ as $N \to \infty$, for $i = 0, 1, \cdots, 2n + 1$.

According to well-known rules for conditional distributions, we have

$$\begin{aligned} P\left[\max_{k/N\varepsilon I_{1/N}} \left| \ln\left(\frac{1}{Z_k^*}\right) - \sum_{l=k}^{N}\frac{1}{l}\right| < \lambda N^{-\frac{1}{2}}\right] &= P\left[\max_{k/N\varepsilon I_{1/N}} \left|\sum_{l=1}^{N+1-k} V_l\right| < \lambda\right] \\ &= \int_{\substack{|x_i|<\lambda \\ i=0,\cdots,2n}} \cdots \int \prod_{\nu=0}^{n} d_{x_{2\nu}} P\left[\max_{l=j_{2\nu}+1,\cdots,j_{2\nu+1}} \left|\sum_{i=1}^{l} V_i\right| < \lambda,\right. \\ &\qquad \left.\sum_{i=1}^{j_{2\nu+1}} V_i \leqq x_{2\nu} \,\middle|\, \sum_{i \leqq j_{2\nu}} V_i = x_{2\nu-1}\right] \\ &\qquad \cdot \prod_{\nu=1}^{n} d_{x_{2\nu-1}} P\left[\sum_{i=1}^{j_{2\nu}} V_i \leqq x_{2\nu-1} \,\middle|\, \sum_{i=1}^{j_{2\nu-1}} V_i = x_{2\nu-2}\right], \end{aligned} \tag{42}$$

where $x_{-1} = 0$. The limits of the probabilities which occur in this integral can be calculated by a limit theorem for partial sums of random variables.

LEMMA. (*See* [13], [14].) *Let* $Y_{M1}, \cdots, Y_{Mm_M}$ *be* m_M *independent random variables with*

$$E(Y_{Mk}) = 0, \qquad \sum_1^{m_M} E(Y_{Mk}^2) = 2t_M.$$

Assume that for all k

$$\frac{E(|Y_{Mk}|^3)}{E(Y_{Mk}^2)} < \mu(M) \tag{43}$$

where $\mu(M) \to 0$ *as* $M \to \infty$, *and let* a, b, ξ, *and* η *be any numbers such that* $a < 0$, $b > 0$, *and* $a \leqq \xi < \eta \leqq b$. *Then*

$$\lim_{M\to\infty} P\left[a < \sum_1^k Y_{Mi} < b, k = 1, \cdots, m_M; \xi < \sum_1^{m_M} Y_{Mi} < \eta\right] = u(0, 0), \tag{44}$$

where $u(s, t)$ *is the solution of the differential equation*

$$\frac{\partial u}{\partial t} = -\frac{\partial^2 u}{\partial s^2} \tag{45}$$

for which the boundary conditions

$$\begin{aligned} u(s, T) &= 0, \quad && a < s < \xi, \qquad \eta < s < b, \\ u(s, T) &= 1, && \xi < s < \eta, \\ u(a, t) &= 0, && 0 < t < T, \\ u(b, t) &= 0, && 0 < t < T, \end{aligned} \tag{46}$$

are fulfilled.

We can apply this lemma to the variables $Y_{j_{2\nu}+1}, Y_{j_{2\nu}+2}, \cdots, Y_{j_{2\nu+1}}$, for $\nu = 0, 1, \cdots, n$, because these variables satisfy (43), with

$$\mu(N) = \frac{3}{f_0 N^{\frac{1}{2}}}.$$

The sum of the second moments

$$2t_M^{(2\nu+1)} = \sum_{j_{2\nu}+1}^{j_{2\nu+1}} \frac{N}{(N + 1 - k)^2} = N \sum_{N+1-j_{2\nu+1}}^{N} \frac{1}{k^2} - N \sum_{N+1-j_{2\nu}}^{N} \frac{1}{k^2}$$

tends towards

$$2T^{(2\nu+1)} = \frac{1 - f_{2n-2\nu}}{f_{2n-2\nu}} - \frac{1 - f_{2n-2\nu+1}}{f_{2n-2\nu+1}} = \frac{f_{2n-2\nu+1} - f_{2n-2\nu}}{f_{2n-2\nu+1}\, f_{2n-2\nu}}, \tag{47}$$

and the boundaries for the partial sums are now

$$a = -\lambda - x_{2\nu-1}, \qquad b = \lambda - x_{2\nu-1}, \qquad \xi = -\lambda - x_{2\nu-1}, \tag{48}$$
$$\eta = x_{2\nu} - x_{2\nu-1},$$

where $x_{-1} = 0$. The solution of (45), which satisfies the boundary conditions (46), (48), is

$$u(s, t) = \frac{1}{2\sqrt{\pi(T^{(2\nu+1)} - t)}} \int_{-\lambda - x_{2\nu-1}}^{x_{2\nu} - x_{2\nu-1}} \sum_{j=-\infty}^{+\infty} (-1)^j \tag{49}$$
$$\cdot \exp\left[-\frac{(s + x_{2\nu-1} - (-1)^j(x + x_{2\nu-1}) - 2\lambda j)^2}{4(T^{(2\nu+1)} - t)}\right] dx.$$

Hence we have

$$\lim_{N\to\infty} P\left[\max_{l=j_{2\nu}+1,\cdots,j_{2\nu+1}} \left|\sum_{i=1}^{l} V_i\right| < \lambda, \sum_{i=1}^{j_{2\nu+1}} V_i \leqq x_{2\nu} \,\middle|\, \sum_{i\leqq j_{2\nu}} V_i = x_{2\nu-1}\right] \tag{50}$$
$$= -\frac{1}{2\sqrt{\pi T^{(2\nu+1)}}} \int_{-\lambda}^{x_{2\nu}} \sum_{j=-\infty}^{+\infty} (-1)^j \exp\left[-\frac{(x - (-1)^j x_{2\nu-1} - 2\lambda j)^2}{4T^{(2\nu+1)}}\right] dx.$$

On the other hand, we apply the Central Limit theorem to the variables $V_{j_{2\nu-1}+1}$, $V_{j_{2\nu-1}+2}, \cdots, V_{j_{2\nu}}$, obtaining

$$\lim_{N\to\infty} P\left[\sum_{i=1}^{j_{2\nu}} V_i \leqq x_{2\nu-1} \,\middle|\, \sum_{i=1}^{j_{2\nu-1}} V_i = x_{2\nu-2}\right] \tag{51}$$
$$= \frac{1}{2\sqrt{\pi T^{(2\nu)}}} \int_{-\infty}^{x_{2\nu-1} - x_{2\nu-2}} e^{-x^2/4T^{(2\nu)}}\, dx,$$

where the $T^{(2\nu)}$ are defined in the same was as the $T^{(2\nu+1)}$ in (47).

In view of (42), (50) and (51) it follows that

$$\lim_{N\to\infty} P\left[\max_{k/N \in I_{1/N}} \left|\sum_{l=1}^{N+1-k} V_l\right| < \lambda\right] = \frac{1}{2^{2n+1}\pi^{n+\frac{1}{2}} \prod_{j=1}^{2n+1} (T^{(j)})^{\frac{1}{2}}}$$
$$\cdot \sum_{p_0,\cdots,p_n=-\infty}^{+\infty} \int \cdots \int_{|x_i|<\lambda} \exp\left[-\sum_{\nu=0}^{n} \frac{(x_{2\nu} - (-1)^{p_\nu} x_{2\nu-1} - 2\lambda p_\nu)^2}{4T^{(2\nu+1)}}\right. \tag{52}$$
$$\left. - \sum_{\nu=0}^{n-1} \frac{(x_{2\nu+1} - x_{2\nu})}{4T^{(2\nu+2)}}\right] dx_0 \cdots dx_{2n},$$

where $x_{-1} = 0$. This expression is $\Psi(\lambda)$. This proves (37) and consequently Theorem 3.

Theorem 4 can be proved in the same way. In the lemma of Kolmogorov we replace a and ξ by $-\infty$. The solution of the boundary problem is now

$$u(s, t) = \frac{1}{2\sqrt{\pi(T^{(2\nu+1)} - t)}} \int_{-\infty}^{x_{2\nu} - x_{2\nu-1}} \sum_{j=0}^{1} (-1)^j \tag{53}$$
$$\cdot \exp\left[-\frac{(s + x_{2\nu-1} - (-1)^j(x + x_{2\nu-1}) - 2\lambda j)^2}{4(T^{(2\nu+1)} - t)}\right] dx,$$

and we obtain

$$
\begin{aligned}
(54)\qquad & \lim_{N\to\infty} P\left[\max_{l=j_{2\nu}+1,\cdots,j_{2\nu+1}}\left(\sum_{i=1}^{l} V_i\right) < \lambda, \left(\sum_{i=1}^{j_{2\nu+1}} V_i\right) \leqq x_{2\nu} \,\middle|\, \left(\sum_{i\leqq j_{2\nu}} V_i\right) = x_{2\nu-1}\right] \\
& = \frac{1}{2\sqrt{\pi T^{(2\nu+1)}}} \int_{-\infty}^{x_{2\nu}} \sum_{j=0}^{1} (-1)^j \exp\left[-\frac{(x-(-1)^j x_{2\nu-1}-2\lambda j)^2}{4T^{(2\nu+1)}}\right] dx.
\end{aligned}
$$

From that Theorem 4 follows.

REFERENCES

[1] T. W. Anderson and D. A. Darling, "Asymptotic theory of certain 'goodness of fit' criteria based on stochastic processes," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 193–212.

[2] Z. W. Birnbaum, "On the power of a one-sided test of fit for continuous probability functions," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 484–489.

[3] Z. W. Birnbaum, "Numerical tabulation of the distribution of Kolmogorov's statistic for finite sample size," *J. Amer. Math. Soc.*, Vol. 47 (1952), pp. 425–441.

[4] Z. W. Birnbaum and F. H. Tingey, "One-sided confidence contours for probability distribution functions," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 592–596.

[5] J. Blackman, "An extension of the Kolmogorov distribution," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 513–520.

[6] J. Blackman, "Correction to 'An extension of the Kolmogorov distribution'," *Ann. Math. Stat.*, Vol. 29 (1958), pp. 318–324.

[7] K. L. Chung, "An estimate concerning the Kolmogorov limit distribution," *Trans. Amer. Math. Soc.*, Vol. 67 (1949), pp. 36–50.

[8] D. A. Darling, "The Kolmogorov-Smirnov, Cramér-von Mises tests," *Ann. Math. Stat.*, Vol. 28 (1957), pp. 823–838.

[9] M. D. Donsker, "Justification and extension of Doobs heuristic approach to the Kolmogorov-Smirnov theorems," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 277–281.

[10] J. L. Doob, "Heuristic approach to the Kolmogorov-Smirnov theorems," *Ann. Math. Stat.*, Vol. 20 (1949), pp. 393–403.

[11] W. Feller, "On the Kolmogorov-Smirnov theorems for empirical distributions," *Ann. Math. Stat.*, Vol. 19 (1948), pp. 177–180.

[12] A. N. Kolmogorov, "Sulla determinazione empirica di una legge di distribuzione," *Giorn. Ist. Ital. Attuari*, Vol. 4 (1943), pp. 83–91.

[13] A. N. Kolmogorov, "Eine Verallgemeinerung des Laplace-Liapounoffschen Satzes," *Bull. Acad. Sci. USSR*, 1931, pp. 956–962.

[14] A. N. Kolmogorov, "Ueber die Grenzwertsätze der Wahrscheinlichkeitsrechnung," *Bull. Acad. Sci. USSR*, 1933, pp. 363–372.

[15] S. Malmquist, "On certain confidence contours for distribution functions," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 523–533.

[16] G. M. Maniya, "Generalization of the criterion of A. N. Kolmogorov," *Doklady Akad. Nauk SSSR*, Vol. 69 (1949), pp. 495–497. (Russian.)

[17] F. J. Massey, Jr., "A note on the power of a non-parametric test," *Ann. Math. Stat.*, Vol. 21 (1950), pp. 440–443.

[18] F. J. Massey, Jr., "A note on the estimation of a distribution function by confidence limits," *Ann. Math. Stat.*, Vol. 21 (1950), pp. 116–119.

[19] A. Rényi, "On the theory of order statistics," *Acta Math. Acad. Sci. Hung.*, Vol. 4 (1953), pp. 191–231.

[20] P. Schmid, "Sur les théorèmes asymptotiques de Kolmogorov et Smirnov pour des fonctions de distributions discontinues," *C. R. Acad. Sci. Paris*, Vol. 243 (1956), pp. 349–351.

[21] N. V. Smirnov, "Sur les écarts de la courbe de distribution empirique," *Mat. Sbornik*, Vol. 6 (1939), pp. 3–26. (Russian, French summary.)

[22] N. V. Smirnov, "Approximation of the distribution law of a random variable by empirical data," *Uspyeki Mat. Nauk SSSR*, Vol. 19 (1944), pp. 179–206. (Russian).

[23] N. V. Smirnov, "Table for estimating the goodness of fit of empirical distributions," *Ann. Math. Stat.*, Vol. 19 (1948), pp. 279–281.

[24] S. S. Wilks, "Order statistics," *Bull. Amer. Math. Soc.*, Vol. 54 (1948), pp. 6–50.

INCOMPLETE SUFFICIENT STATISTICS AND SIMILAR TESTS[1]

By Robert A. Wijsman[2]

University of California, Berkeley

0. Summary. For a family of exponential densities a method is given, called "D method," for constructing a class of similar tests in the case that the minimal sufficient statistic is boundedly incomplete. This method also provides a proof of a criterion for bounded incompleteness. Under certain conditions the criterion states that a sufficient statistic for a family of exponential densities is boundedly incomplete if the number of components of the statistic is larger than the number of parameters specifying the distribution. Applications are indicated in the Behrens-Fisher problem, and in the problem of testing the ratio of mean to standard deviation in a normal population. In the latter problem it is shown that the D method generates the whole class of similar tests. Some unsolved problems concerning the existence of an optimal similar test are indicated.

1. Introduction. Lehmann and Scheffé [8], [9] have introduced the concept of completeness of a family of measures and have shown the usefulness of this notion both for unbiased estimation and for the construction of similar regions. The latter were introduced by Neyman and Pearson [11] as a means to cope with tests of composite hypotheses. If the hypothesis is composite only because of nuisance parameters, then the requirement of similarity of the test is often a convenient means of restricting the class of tests to be considered. If the hypothesis is composite both of nuisance parameters and because the parameter tested is not completely specified by the hypothesis, then similarity is often required if the test is to be unbiased. For instance, let θ be a real parameter, τ a possibly vector valued nuisance parameter, and let the hypothesis be $H:\theta \leqq \theta_0$, the alternative $\bar{H}:\theta > \theta_0$, for some specified θ_0. Suppose we want the test to be unbiased, then the power function of the test has to be $\leqq \alpha$ for $\theta \leqq \theta_0$ and $\geqq \alpha$ for $\theta > \theta_0$, where α is the level of significance. If, in addition, the power function is continuous, which is usually the case, then we have automatically that its value on the surface $\theta = \theta_0$ equals α, identically in τ. Search for an optimum unbiased test reduces then to the simpler problem of search for an optimum similar test of the hypothesis $H_1:\theta = \theta_0$ against $\bar{H}:\theta > \theta_0$.

In the presence of a sufficient statistic there exists a special class of easily constructible similar regions [10], termed *similar regions of Neyman structure* by Lehmann and Scheffé [8]. They proved that every similar region is of Neyman structure if and only if the family of distributions of the sufficient statistic, as specified by the hypothesis, is boundedly complete [8]. Unfortunately, there

Received August 29, 1957; revised June 5, 1958.

[1] This investigation was supported (in part) by a research grant (No. G-3666) from the National Institutes of Health, Public Health Service.

[2] Now at the University of Illinois, Department of Mathematics.

are important problems in which the latter condition is not fulfilled, in which case the class of all similar regions is larger than the class of similar regions of Neyman structure. An example is the Behrens-Fisher problem (see, for example, [13], in which also references to earlier work can be found). In this problem the similar regions of Neyman structure are of no use, since for any such region the power function is identically constant.

All remarks in the previous paragraph are equally valid if instead of similar rejection regions we consider randomized similar tests. It is clear from the discussion that in each problem of testing a composite hypothesis by means of a similar test it is important to know whether or not the problem admits a boundedly complete sufficient statistic. If not, one would like to have a method of constructing all similar tests. It is the purpose of this paper to provide partial answers to these problems. In section 3 a method termed the "D method," will be given for the construction of a large class of similar tests in the case of a family of exponential densities. In section 5 the D method will be used to derive a criterion for bounded incompleteness in the case of a family of exponential densities. Two examples of the D method are given in section 4; the first example is the Behrens-Fisher problem, the second example is the problem of testing the ratio of mean to standard deviation in a normal population. For the latter problem it is proved in section 6 that every similar test can be constructed by the D method, provided this method is given sufficiently wide scope. Some remarks on the problem of finding an optimal similar test are made in section 7. A preliminary account of the results of sections 3 and 5 appeared in [16].

2. Similar tests and boundedly incomplete sufficient statistics. Let $\mathfrak{X}$ be a space of points x, $\mathfrak{A}$ a σ-field of subsets of $\mathfrak{X}$ (with $\mathfrak{X} \varepsilon \mathfrak{A}$), and $\mathcal{P} = \{P_\theta, \theta \varepsilon \Omega\}$ a family of probability measures on $(\mathfrak{X}, \mathfrak{A})$. Expectation with respect to P_θ will be denoted by E_θ. If $\omega \subset \Omega$ and T is a sufficient statistic for $\mathcal{P}_\omega = \{P_\theta, \theta \varepsilon \omega\}$, we shall also say that T is a sufficient statistic for ω. The range of T is denoted by $\mathfrak{I}$, and is understood to be a Borel subset of a Euclidean space. Let $\mathfrak{B}$ be the σ-field of Borel subsets of $\mathfrak{I}$. We recall the following definitions: A sufficient statistic for ω is called *minimal* if the sufficient sub σ-field which it induces in $\mathfrak{X}$ is "essentially" contained in every sufficient sub σ-field for ω (see Bahadur [2] for a precise definition).[3] A sufficient statistic T for ω is called *complete* for ω if, for every $\mathfrak{B}$-measurable numerical function g

$$(1) \qquad E_\theta g(T) = 0 \quad \text{for all} \quad \theta \varepsilon \omega \Rightarrow g \equiv 0 \quad \text{a.e.} \quad (\mathcal{P}_\omega).$$

If the implication (1) holds for every *bounded* $\mathfrak{B}$-measurable numerical function, then T is called *boundedly complete* for ω. The following implications are true [8].

$$(2) \qquad \text{Completeness} \Rightarrow \text{bounded completeness} \Rightarrow \text{minimality}.$$

[3] The term *minimal* was introduced by Lehmann and Scheffé [8], whereas Bahadur [2] describes the same concept with the term *necessary and sufficient* statistic.

Suppose a composite hypothesis H specifies $\theta \varepsilon \omega \subset \Omega$. We shall consider randomized tests for H with test functions ϕ, where, for each $x \varepsilon \mathfrak{X}$, $0 \leqq \phi(x) \leqq 1$, ϕ measurable, and H is rejected with probability $\phi(x)$ if x is observed. Among all tests we restrict ourselves to similar tests, defined by the condition that $E_\theta\phi$ is independent of θ if $\theta \varepsilon \omega$. If T is a sufficient statistic for ω, $\mathfrak{A}_0 \subset \mathfrak{A}$ its sufficient sub σ-field and ϕ any test, we can consider the $\mathfrak{A}_0$-measurable function $E(\phi \mid \mathfrak{A}_0)$. If α is a number, $0 < \alpha < 1$, and if ϕ is such that $E(\phi \mid \mathfrak{A}_0) \equiv \alpha$, then clearly $E_\theta\phi = \alpha$ for all $\theta \varepsilon \omega$, so that ϕ is similar. Such a ϕ is called a *test of Neyman structure* [8]. If T is a boundedly complete sufficient statistic for ω, then every similar test has Neyman structure [8]. On the other hand, if a sufficient statistic T is not boundedly complete for ω, then there exist similar tests which do not have Neyman structure. This follows from the fact that the bounded incompleteness implies the existence of a $\mathfrak{B}$-measurable numerical function g on $\mathfrak{J}$, bounded below by $-\alpha$, above by $1 - \alpha$, different from 0 on a set of positive probability (with respect to $\mathcal{P}_\omega$), with $E_\theta g(T) = 0$ for all $\theta \varepsilon \omega$. With f on $\mathfrak{X}$ defined by $f(x) = g(T(x))$, we have that $\phi = f + \alpha$ is similar of size α, but $E(\phi \mid \mathfrak{A}_0) - \alpha = f \not\equiv 0$ on a set of positive probability, so ϕ is not a test of Neyman structure. Conversely, for any similar test ϕ we can form the function $f = E(\phi \mid \mathfrak{A}_0) - \alpha$ and define g on $\mathfrak{J}$ by $g(T(x)) = f(x)$, so that

$$E_\theta g(T) = 0$$

for all $\theta \varepsilon \omega$. It follows that all similar tests can be found by constructing all bounded numerical functions g on $\mathfrak{J}$ whose expectations vanish for all $\theta \varepsilon \omega$.

3. The D method for constructing similar tests in the case of a family of regular exponential densities. In this section the restriction of θ to ω will be understood. Let the distribution of T, induced by P_θ, have a density with respect to m-dimensional Lebesgue measure, and let this density p_θ be of the form

$$p_\theta(t) = c(\theta) \exp\left[-\sum_{i=1}^{m} s_i(\theta)t_i\right] h(t) \tag{3}$$

in which $t = (t_1, \cdots, t_m)$, and $s_1, \cdots, s_m$ are real valued functions on ω. We shall assume that the function h is of such a nature that it is possible to find a closed m-dimensional cube C on which h is bounded away from 0. With this restriction on h, the family (3) will be called a family of *regular exponential densities*. Exponential densities which arise in statistics are always regular.

If ω is an m-dimensional subset of an m-dimensional Euclidean space, then, under mild conditions, T with density (3) is complete for ω [9]. In that case every similar test has Neyman structure. From the point of view of the present paper the interesting case arises when ω is a subset of an $m - 1$ dimensional Euclidean space. In that case θ has at most $m - 1$ components, so that the m functions s_i are functions of at most $m - 1$ parameters. Eliminating those parameters will result in a functional relation between the s_i. Suppose that this relation can be put in the form

$$P(s_1, \cdots, s_m) = 0 \tag{4}$$

in which P is a polynomial of positive degree in at least one of the s_i. It should be kept in mind that (4) holds identically in θ.

As discussed in section 2, a similar test of non-Neyman structure can be constructed by constructing a bounded function g on $\mathfrak{J}$, $g \not\equiv 0$ on a set of positive probability, such that

$$\int g(t)p_\theta(t)\,dt \underset{\theta}{\equiv} 0 \tag{5}$$

Using (3), remembering that h is bounded away from 0 on some m-dimensional cube C, it suffices to construct a bounded function F which is $\not\equiv 0$ on a subset of C of positive Lebesgue measure, vanishes outside C, and satisfies

$$\int F(t) \exp\left[-\sum_i^m s_i(\theta)t_i\right] dt \underset{\theta}{\equiv} 0 \tag{6}$$

The function g in (5) can then be taken as F/h. The left hand side of (6) is the m-dimensional Laplace transform of F, denoted by $\mathfrak{L}(F)$:

$$\int F(t_1 \cdots, t_m) \exp\left[-\sum_i^m s_i\, t_i\right] dt = \mathfrak{L}(F)(s_1 \cdots, s_m). \tag{7}$$

The problem is to construct F in such a way that $\mathfrak{L}(F) = 0$ for all values of $s(\theta)$, $\theta \in \omega$. This can be done with help of (4). Let P be of degree d and let G be a function on $\mathfrak{J}$ possessing all partial derivatives of dth order in the interior of C, vanishing outside C, and having all partial derivatives of $d - 1$st order continuous on the boundary of C. An example of such a function is the following. Let C be given by $a_i \leqq t_i \leqq a_i + l$ $(i = 1, \cdots, m)$, then on C we can take $G(t) = \prod_{i=1}^m (t_i - a_i)^d(a_i + l - t_i)^d$. Now denote by D the differential operator

$$D = P\left(\frac{\partial}{\partial t_1}, \cdots, \frac{\partial}{\partial t_m}\right). \tag{8}$$

We then have

$$\mathfrak{L}(DG)(s) = P(s)\mathfrak{L}(G)(s) \tag{9}$$

in which $s = (s_1, \cdots, s_m)$. Since the right hand side of (9) is $\equiv 0$ by (4), we may take F in (7) to be $F = DG$. The final result is therefore

$$g(t) = (DG(t))/h(t) \tag{10}$$

for suitably chosen G, and

$$\phi(t) = \alpha + (DG(t))/h(t) \tag{11}$$

is a size α similar test of non-Neyman structure.

Even for one m-dimensional cube C the number of choices for G is large. In addition there will usually be a large number of m-dimensional cubes on each of which h is bounded away from 0, and finally one may consider regions other than cubes for which the construction of functions G is possible. Thus, there will be a large class of functions g satisfying (5) which can be generated by the

differential operator method, called the *D method* henceforth. Whether this method, in general, will give all those functions g, is still an open question. In one particular case the question has been answered in the affirmative, provided the definition of D method is taken sufficiently wide (see section 6).

Suppose that with the help of the D method a similar test $\phi(T)$ is constructed, and that it is desired to consider similar tests which do not necessarily depend on T only. Let ψ be a test function defined on the sample space $\mathfrak{X}$. If ψ is chosen to satisfy $E(\psi \mid t) = \phi(t)$, then ψ is also similar. In particular, it will usually be possible to construct in this way a similar rejection region w, in which case ψ is the indicator of w (this construction fails is $\mathfrak{X}$ is a subspace of a Euclidean space with same dimension as $\mathfrak{J}$) A similar region w is constructed by demanding

$$P(w \mid t) = \phi(t). \tag{12}$$

In other words, on each surface $T = t$ in the sample space a region is selected which has conditional probability $\phi(t)$. This generalizes the construction of a similar region of Neyman structure [10]. Equation (12) will be used in section 4, example 2.

4. Examples of the D method. EXAMPLE 1 (Behrens-Fisher problem). Let $X_1, \cdots, X_{n_1}$ be n_1 independent observations on a normal variable with mean μ_1, variance σ_1^2, and $Y_1, \cdots, Y_{n_2}$, n_2 independent observations on a normal variable with mean μ_2, variance σ_2^2. The X's and Y's are independent, and all parameters are unknown. Under the hypothesis tested, which is $\mu_1 = \mu_2$, the joint distribution of the X's and Y's has an exponential density with exponential factor

$$\exp\left[-\frac{1}{2\sigma_1^2}\sum_1^{n_1} x_i^2 + \frac{\mu}{\sigma_1^2}\sum_1^{n_1} x_i - \frac{1}{2\sigma_2^2}\sum_1^{n_2} y_j^2 + \frac{\mu}{\sigma_2^2}\sum_1^{n_2} y_j\right]$$

in which μ is the common value of μ_1 and μ_2. We may take

$$T_1(x) = \sum_1^{n_1} x_i^2, \qquad T_2(x) = \sum_1^{n_1} x_i, \qquad T_3(x) = \sum_1^{n_2} y_j^2, \qquad T_4(x) = \sum_1^{n_2} y_j,$$

$$s_1(\theta) = \frac{1}{2\sigma_1^2}, \qquad s_2(\theta) = \frac{-\mu}{\sigma_1^2}, \qquad s_3(\theta) = \frac{1}{2\sigma_2^2}, \qquad s_4(\theta) = \frac{-\mu}{\sigma_2^2}.$$

The s_i are linearly independent, from which it can be shown that

$$T = (T_1, T_2, T_3, T_4)$$

is a minimal sufficient statistic for ω. T has a regular exponential density of form (3), with

$$h(t) = (n_1t_1 - t_2^2)^{(n_1-3)/2}(n_2t_3 - t_4^2)^{(n_2-3)/2} \tag{13}$$

if $n_1t_1 \geqq t_2^2$, $n_2t_3 \geqq t_4^2$, and $h(t) = 0$ otherwise. By eliminating μ, σ_1, σ_2 from

the four s_i we obtain $s_1 s_4 - s_2 s_3 = 0$ as a realization of (4). The differential operator D in (8) is then

$$D = \frac{\partial}{\partial t_1}\frac{\partial}{\partial t_4} - \frac{\partial}{\partial t_2}\frac{\partial}{\partial t_3} \tag{14}$$

and for suitably chosen G the test

$$\phi(t) = \alpha + h^{-1}(t)\left(\frac{\partial}{\partial t_1}\frac{\partial}{\partial t_4} - \frac{\partial}{\partial t_2}\frac{\partial}{\partial t_3}\right)G(t) \tag{15}$$

is similar and of size α, where $h(t)$ is given by (13). Whether this method can be used to show the existence of an invariant similar region, such as the one proposed by Welch [1], [15], has not yet been investigated.

It should perhaps be mentioned here that the approach to the Behrens-Fisher problem by Wald [14] is essentially different, since Wald does not require the test to be similar.

Example 2. (Standardized mean of a normal population). Suppose we make $n + 1$ independent observations on a normal variable and consider hypotheses concerning the ratio of mean to standard deviation. By an orthogonal transformation this problem can be brought in the following form: Let $X_0, \cdots, X_n$ be independent and normal, with common, unknown variance σ^2. X_0 has unknown mean μ, $X_1, \cdots, X_n$ have mean 0. Denote $\mu/\sigma = r$, then for some given r_0 the hypothesis tested is $r = r_0$. For the time being the alternative to be considered is immaterial. For later reference, however, suppose that the alternative is $r > r_0$. We then have

$$\Omega = \{(r, \sigma): r \geqq r_0, \sigma > 0\},$$

$$\omega = \{(r, \sigma): r = r_0, \sigma > 0\}.$$

Under the hypothesis the joint distribution of the X_i has the form given by (3), with exponential factor

$$\exp\left[-\frac{1}{2\sigma^2}\sum_0^n x_i^2 + \frac{r_0}{\sigma}x_0\right]$$

so that we may take

$$T_1(x) = \sum_0^n x_i^2, \qquad T_2(x) = x_0, \qquad s_1(\sigma) = \frac{1}{2\sigma^2}, \qquad s_2(\sigma) = -\frac{r_0}{\sigma}.$$

$T = (T_1, T_2)$ is minimal sufficient, since s_1 and s_2 are linearly independent. Elimination of σ from s_1 and s_2 gives $s_2^2 - 2r_0^2 s_1 = 0$, so that we can take

$$P(s_1, s_2) = s_2^2 - 2r_0^2 s_1 \tag{16}$$

and

$$D = \frac{\partial^2}{\partial t_2^2} - 2r_0^2\frac{\partial}{\partial t_1} \tag{17}$$

The function h in (3) is found to be

$$h(t_1, t_2) = (t_1 - t_2^2)^{(n/2)-1} \tag{18}$$

if $t_1 \geqq t_2^2$, and $h = 0$ otherwise. For suitably chosen $G(t_1, t_2)$ the test function

$$\phi(t_1, t_2) = \alpha + (t_1 - t_2^2)^{1-n/2} \left(\frac{\partial^2}{\partial t_2^2} - 2r_0^2 \frac{\partial}{\partial t_1}\right) G(t_1, t_2) \tag{19}$$

is similar and of size α.

Equation (19) can be used to demonstrate the existence of similar tests which are not invariant. In the present problem an invariant test is a function of $T_2/\sqrt{T_1}$ only. Choose for G in (19) the following function:

$$G(t_1, t_2) = c(t_1 - t_2^2)^{(n/2)+1} e^{-t_1} \tag{20}$$

if $t_1 \geqq t_2^2$ and $G = 0$ otherwise, with $c > 0$ chosen so small that ϕ is bounded between 0 and 1. It is easily checked that after substitution into (19) the resulting test function is not invariant. This example can also be used to show the existence of similar rejection regions which are not equivalent to a cone in the sample space (we shall call two tests *equivalent* if they have the same power function, and by a *cone* is meant a union of rays through the origin). If w is any rejection region, ϕ the corresponding test function, given by (12), then w and ϕ are equivalent since T is sufficient, not only for ω, but also for Ω. If w_1 is a cone in the sample space, then the corresponding ϕ_1 is invariant. Let w_2 be any rejection region equivalent to w_1, ϕ_2 the corresponding test function; then ϕ_1 and ϕ_2 are equivalent. Now T is not only sufficient for Ω, it is also complete for Ω. Since ϕ_1 and ϕ_2 have the same power functions, it follows then that $\phi_1 = \phi_2$ a.e. and thus ϕ_2 is also invariant. The existence of a noninvariant similar test ϕ implies then the existence of a similar region which is not equivalent to any cone in the sample space.[4]

5. A criterion for bounded incompleteness in the case of regular exponential densities. Let the family of distributions be given by (3), with $\theta \, \varepsilon \, \omega$. By (2), if T is not minimal sufficient for ω, then T cannot be boundedly complete. This happens, for instance, if the s_i are linearly dependent on ω because the exponent $-\Sigma s_i t_i$ in (3) can then be written as a linear combination of fewer than m of the t_i. The incompleteness in this case also follows from the applicability of the D method of section 3, because of the existence of a polynomial P, linear in this case, for which (4) holds. On the other hand, if the m functions s_i are linearly independent on ω, then T is minimal sufficient for ω. Even if this is the case, T may still be boundedly incomplete. Theorem 2 below tells when this will happen. Its proof uses the D method of section 3. The conditions of Theorem 2 are designed to guarantee the existence of the polynomial P on the left

[4] This seems to contradict a statement by Patnaik [12] to the effect that in the problem under consideration every similar region is equivalent to a cone in the sample space. However, Patnaik's proof is unconvincing, and the non-invariant ϕ exhibited above provides a counter example.

hand side of (4), such that P is not the zero polynomial. This is made possible by the following theorem, due to A. Seidenberg (private communication). The proof is given in Appendix 1.

THEOREM 1. (Seidenberg). *Let for each i, $i = 1, \cdots, m$, $P_i(s_i; \theta_1, \cdots, \theta_k)$ be a polynomial in s_i and the $\theta_j (j = 1, \cdots, k)$, with coefficients in some field K, where $k < m$ and P_i is of positive degree in s_i. Let $A_i(\theta)$ be the leading coefficient of P_i as a polynomial in s_i. Then there is a polynomial $P(s_1, \cdots, s_m)$ with coefficients in K, which is not the zero polynomial, and a power product $B(\theta)$ of the $A_i(\theta)$, such that $B(\theta)P(s) = 0$ whenever $P_i = 0$ for all i.*

COROLLARY. *If θ is restricted to a set Θ, and if, for each $\theta \in \Theta$ and each i,*

$$A_i(\theta) \neq 0,$$

then $P = 0$ whenever $P_i = 0$ for all i.

For, if $A_i(\theta) \neq 0$, $i = 1, \cdots, m$, then $B(\theta) \neq 0$.

In the application we want to make of the corollary, the set Θ is ω. Furthermore, we shall assume the s_i of section 3 to be algebraic functions of the θ_j, for $\theta \in \omega$. Then for each i there is a polynomial P_i in s_i and the θ_j such that $P_i(s_i; \theta_1, \cdots, \theta_k) = 0$ if $\theta \in \omega$. We shall further assume that the $A_i(\theta)$ are $\neq 0$ if $\theta \in \omega$. These conditions will be satisfied in particular if, for each i, s_i on ω is a rational function of the θ_j, with nonvanishing denominator.

THEOREM 2. *Suppose a family of regular exponential densities is given by* (3), *with $\theta \in \omega$; ω is a subset of a k-dimensional Euclidean space, with $k < m$; on ω, the m functions s_i are algebraic functions of the k parameters θ_j, so that*

$$P_i(s_i; \theta_1, \cdots, \theta_k) = 0$$

for some polynomial P_i ($i = 1, \cdots, m$); $A_i(\theta)$, the leading coefficient of P_i as a polynomial in s_i, does not vanish anywhere on ω for any i. Then T is boundedly incomplete for ω.

The proof follows immediately from the constructibility, by the D method of section 3, of a bounded function g, $g \neq 0$ on a set of positive probability, satisfying $E_\theta g(T) = 0$ for all $\theta \in \omega$.

In both examples in section 4 the s_i are rational functions of the θ_j, with nonvanishing denominators, and in both cases $k = m - 1 < m$, so that Theorem 2 applies. This provides another proof of the well-known fact that in the Behrens-Fisher problem, as well as in the problem of testing the ratio of mean to standard deviation in a normal population, the minimal sufficient statistic is boundedly incomplete.

It would be interesting to know how much the assumptions of Theorem 2 can be relaxed. It is certainly not necessary that the s_i be algebraic functions of the θ_j, for, if $m = 2$, $k = 1$, $s_1 = \cos\theta$, $s_2 = \sin\theta$, then $s_1^2 + s_2^2 - 1 = 0$, as a realization of (4), so that the D method applies. It is not even necessary for incompleteness that there exists a polynomial P in the s_i which vanishes for all $\theta \in \omega$, as the next example will show. Take $m = 2$, $k = 1$, $s_1 = -\ln\theta$, $s_2 = -\ln(1 - \theta)$, with $0 < \theta < 1$. Instead of (4) we have a transcendental equation:

$\exp[-s_1] + \exp[-s_2] - 1 = 0$. With help of this equation one can easily construct functions F of the kind mentioned in section 3. For example, the function F whose 2-dimensional Laplace transform is

$$\mathfrak{L}(F)(s_1, s_2) = \frac{1}{s_1 s_2}(e^{-s_1} + e^{-s_2} - 1)(e^{-a_1 s_1} - e^{-b_1 s_1})(e^{-a_2 s_2} - e^{-b_2 s_2})$$

is bounded between -1 and 1, vanishes outside the rectangle

$$a_1 \leqq t_1 \leqq b_1 + 1, \qquad a_2 \leqq t_2 \leqq b_2 + 1,$$

and has vanishing Laplace transform for all θ between 0 and 1. On the other hand, the fact that $k < m$ is not sufficient for bounded incompleteness, nor is the additional restriction of analyticity of the s_i sufficient. The following example is due to L. J. Savage (private communication). In (3) choose $m = 2$, $k = 1$, $s_1 = \theta \cos \theta$, $s_2 = \theta \sin \theta (\theta > 0)$, $h(t) = 1$ for t in some square, $h = 0$ otherwise. Here the s_i are analytic functions of θ, but yet it can be shown that the family of distributions is complete. Another example is due to D. L. Burkholder (private communication) and differs from Savage's example only in that $s_1 = \theta \cos (1/\theta)$, $s_2 = \theta \sin (1/\theta)$. This example is a little less regular than Savage's example, but on the other hand the completeness of the family of distributions is easier to show.

6. Completeness of the D method in the case of a hypothesis concerning the standardized mean of a normal population. In this section it will be shown that in Example 2 of section 4 all similar tests can be generated by the D method, provided the D method is defined in a sufficiently broad manner. That is, we want to show that for each similar test ϕ there exists a function G satisfying (19) and certain other conditions. In section 3 the functions G were restricted to some m-dimensional cube on which h is bounded away from 0 but it was remarked there that this restriction is not necessary. We shall not even demand that $G = 0$ whenever $h = 0$. In fact, the main thing of importance was the validity of (9), and even this we shall relax slightly in the problem under consideration.

Equation (19) can be put in the form

$$\left(\frac{\partial}{\partial t_1} - \frac{1}{2r_0^2}\frac{\partial^2}{\partial t_2^2}\right) G = \frac{\sqrt{2\pi}}{r_0}\varphi \tag{21}$$

where φ is defined by

$$\varphi(t) = -(\sqrt{8\pi}r_0)^{-1}h(t)(\phi(t) - \alpha). \tag{22}$$

Equation (21) can be considered as the heat equation in one dimension, if t_1 is interpreted as time, t_2 as position, G as temperature, and $(\sqrt{2\pi}/r_0)\varphi$ as a heat source, capable of producing both positive and negative heat, whose strength and spatial distribution varies with time. If this were an actual heat problem, its solution could be written down at once, employing the usual Green's function for the heat operator:

$$G(t_1, t_2) = \int\int \varphi(t_1', t_2')(t_1 - t_1')^{-1/2} \exp\left[-\frac{r_0^2}{2}\frac{(t_2 - t_2')^2}{t_1 - t_1'}\right] dt_1' \, dt_2' \tag{23}$$

where the integration is over the strip $0 \leqq t_1' \leqq t_1$. Since $h(t')$, and therefore, $\varphi(t')$, is zero unless $t_2'^2 \leqq t_1'$, we may integrate over $t_2'^2 \leqq t_1' \leqq t_1$. The question to be answered next is whether, and if so, in what sense, the formal solution (23) to (21), and therefore, to (19) is a representation of ϕ.

We shall at once study the power function of any similar test ϕ, since some of the results are needed in section 7. Let Ω and ω be as defined in section 4, Example 2. We shall assume $r_0 > 0$. As remarked in section 4, the statistic

$$T = (T_1, T_2)$$

is sufficient for Ω, and it suffices therefore to consider test functions ϕ which depend only on T. The power function of ϕ is $\beta(r, \sigma) = E_{r,\sigma}\phi(T_1, T_2)$. Suppose ϕ satisfies (19), then we get after substitution:

$$\begin{aligned} \beta(r, \sigma) &= \alpha + c(r, \sigma) \int\int \\ &\cdot \exp\left[-\frac{1}{2\sigma^2}t_1 + \frac{r}{\sigma}t_2\right]\left(\frac{\partial^2}{\partial t_2^2} - 2r_0^2\frac{\partial}{\partial t_1}\right) G(t_1, t_2) \, dt_1 \, dt_2 \end{aligned} \tag{24}$$

where the integration is over $0 \leqq t_1 < \infty$, $-\infty < t_2 < \infty$. We may effect this integration by taking the upper limits on t_1 and t_2 as A, B respectively, and then let $A \to \infty$, $B \to \infty$ in any order. With respect to the types of functions G to be considered it will not be necessary to do something similar with the lower limit on t_2. If the upper limits on t_1 and t_2 are A and B, one can integrate by parts, obtaining an integral

$$\frac{r^2 - r_0^2}{\sigma^2}\int_0^A dt_1 \int_{-\infty}^B G(t_1, t_2) \exp\left[-\frac{1}{2\sigma^2}t_1 + \frac{r}{\sigma}t_2\right] dt_2 \tag{25}$$

plus the following integrated terms:

$$-2r_0^2 \int_{-\infty}^B G(A, t_2) \exp\left[-\frac{1}{2\sigma^2}A + \frac{r}{\sigma}t_2\right] dt_2 \tag{26}$$

$$-\frac{r}{\sigma}\int_0^A G(t_1, B) \exp\left[-\frac{1}{2\sigma^2}t_1 + \frac{r}{\sigma}B\right] dt_1 \tag{27}$$

$$\int_0^A \frac{\partial G(t_1, B)}{\partial t_2} \exp\left[-\frac{1}{2\sigma^2}t_1 + \frac{r}{\sigma}B\right] dt_1 \tag{28}$$

There is also an integral involving G on the t_2-axis. For any G given by (23), $G(0, t_2) = 0$, so that the integral mentioned in the preceding sentence vanishes trivially. It is sufficient, then, to consider only functions G which vanish if $t_1 = 0$. Now if G is given by (23), with φ defined by (22) and ϕ similar of size α, then it can be shown that (26)–(28) vanish in the limit if we let first $B \to \infty$ and then $A \to \infty$. A proof is given in Appendix 2. Using (24) and (25) it follows that

$$\beta(r, \sigma) = \alpha + c(r, \sigma) \frac{r^2 - r_0^2}{\sigma^2} \lim_{A\to\infty} \lim_{B\to\infty} \int_0^A dt_1 \int_{-\infty}^B G(t_1, t_2) \cdot \exp\left[-\frac{1}{2\sigma^2} t_1 + \frac{r}{\sigma} t_2\right] dt_2 \tag{29}$$

We see from (29) that $\beta(r_0, \sigma) = \alpha$ identically in σ, as it should.

The reason we could get the power function in the form (29) is that in this problem the density of T is of the exponential form (3) on the whole of Ω. The exponent of the exponential factor is $-t_1/(2\sigma^2) + rt_2/\sigma$, so that on Ω we have $s_1 = 1/(2\sigma^2)$, $s_2 = -r/\sigma$. The polynomial (16) is now defined on the whole of Ω:

$$P(s) = s_2^2 - 2r_0^2 s_1 = \frac{r^2 - r_0^2}{\sigma^2} \tag{30}$$

(On ω, $r = r_0$, so $P = 0$ as it should). We made the integrated terms (26)–(28) vanish by taking limits in a special way. This suggests, for this problem to redefine the 2-dimensional Laplace transform as follows:

$$\mathfrak{L}(F)(s_1, s_2) = \lim_{A\to\infty} \lim_{B\to\infty} \int_0^A dt_1 \int_{-\infty}^B F(t_1, t_2) \exp\left[-s_1 t_1 - s_2 t_2\right] dt_2 \tag{31}$$

With P and $\mathfrak{L}$ defined by (30) and (31), we have proved that if ϕ is similar, and G is the corresponding function given by (23), then (9) is valid on the whole of Ω. Adding α to both sides of (9) then produces (29).

In order to characterize the whole class of similar tests, consider the class $\mathcal{C}$ of functions G defined on the right half (t_1, t_2) plane which satisfy the following conditions (with D defined by (17)):

(i) $DG(t_1, t_2) = 0$ if $t_2^2 > t_1$,

(ii) $-\alpha \leqq (t_1 - t_2^2)^{1-n/2} DG(t_1, t_2) \leqq 1 - \alpha$ if $t_2^2 \leqq t_1$,

(iii) $G = 0$ if $t_1 = 0$, and $G(t_1, t_2) \to 0$ as $t_2 \to -\infty$, for each t_1,

(iv) The integrals (26)–(28) approach 0 if we let first $B \to \infty$ and then $A \to \infty$.

For every similar size α test function ϕ there is, by (23) and (22), a unique G, satisfying the conditions (i)–(iv), so that $G \in \mathcal{C}$. Conversely, for any $G \in \mathcal{C}$ we have shown that ϕ given by (19) is similar and of size α. Thus, there is a one-to-one correspondence between the members of $\mathcal{C}$ and the similar size α test functions. The class $\mathcal{C}$ gives therefore a complete characterization of the similar tests. Unfortunately, condition (iv) is not a very easy one. There is an important subclass of $\mathcal{C}$ where (iv) is obviously fulfilled, consisting of those functions G in $\mathcal{C}$ which vanish identically for $t_2 > \sqrt{t_1}$. This is the case, for instance, with all functions G leading to an invariant test. For a proof of this fact see Appendix 3. It would be desirable if (iv) could be replaced by a simpler condition. The possibility is not excluded that conditions (i)–(iv) imply that $G(t_1, t_2) = 0$ for all $t_2 > \sqrt{t_1}$, but whether this is so is an open question.

7. Some remarks on the search for an optimum test in the problem of section 6. Consider the class $\mathcal{C}$ defined in section 6. Let ϕ_1, ϕ_2 be two similar size α tests, G_1, G_2 the corresponding functions in $\mathcal{C}$, and β_1, β_2 their power functions. It follows from (29), since $r^2 \geqq r_0^2$, that if $G_1 \geqq G_2$, then $\beta_1 \geqq \beta_2$, so that ϕ_1 is uniformly more powerful than ϕ_2. Since every similar ϕ has a representative $G \varepsilon \mathcal{C}$, if there would exist a $G_0 \varepsilon \mathcal{C}$ such that $G_0 \geqq G$ for every $G \varepsilon \mathcal{C}$, then the test function ϕ_0 corresponding to G_0 would be UMP (uniformly most powerful) among all similar tests. To decide whether or not such a dominating function G_0 exists, the following observations may be of help. The first observation is that in the problem under consideration every invariant test—that is, depending only on $T_2/\sqrt{T_1}$—is similar. Secondly, if we denote by $\mathcal{C}^*$ the subclass of $\mathcal{C}$ representing invariant tests, then in $\mathcal{C}^*$ there is a function G_0^* which dominates every $G^* \varepsilon \mathcal{C}^*$. The corresponding test function ϕ_0^* is therefore UMP among all invariant tests. ϕ_0^* is nonrandomized, with a rejection region of the form $t_2/\sqrt{t_1} >$ constant. That ϕ_0^* is UMP invariant is a known result [12], obtainable more directly by the observation that $T_2/\sqrt{T_1 - T_2^2}$ has a noncentral t-distribution with a monotonic likelihood ratio [3], [4], [7]. The third observation we want to make is that if the dominating function G_0 exists, it has to coincide with G_0^*. This follows from the following proposition: *If a* UMP *similar test based on T exists, it is necessarily invariant.* The analogous statement, with "similar" replaced by "unbiased," is well known [5], [6]. In fact, both statements are special cases of the following more general theorem, due to E. L. Lehmann (private communication): *Let $\mathcal{G}$ be a group of transformations which leaves the problem invariant, and let $\mathcal{K}$ be a class of tests which is closed under $\mathcal{G}$. If there is a unique* UMP *test in $\mathcal{K}$, it is almost invariant.* (The uniqueness is understood to be a.e.). The proof of this theorem follows the same lines as in the special case that $\mathcal{K}$ is the class of unbiased tests of fixed size. In our problem $\mathcal{K}$ is the class of similar tests of size α, based on T. $\mathcal{K}$ is clearly closed under $\mathcal{G}$. If there is a UMP test in $\mathcal{K}$, its uniqueness follows from the completeness of T for Ω. Finally, in our problem an almost invariant function can be shown to be invariant (see also [17], footnote 3, and [5]).

The conclusion drawn from the preceding discussion is that there is a dominating function $G_0 \varepsilon \mathcal{C}$ if and only if G_0^* is the dominating function. Whether or not this is so is still an open question, and consequently, it is still unknown whether a UMP similar test exists. A last remark may be added to this. As remarked in section 6 and proved in Appendix 3, the functions G^* in $\mathcal{C}^*$ have the remarkable property that they vanish for $t_2 \geqq \sqrt{t_1}$. This property holds then in particular for G_0^*. Taking into account that $G \varepsilon \mathcal{C} \Rightarrow -aG \varepsilon \mathcal{C}$ for sufficiently small $a > 0$, we conclude that if G_0^* is a dominating function in $\mathcal{C}$, then every $G \varepsilon \mathcal{C}$ must also have the property $G(t_1, t_2) = 0$ if $t_2 \geqq \sqrt{t_1}$. If this were indeed true, then condition (iv) in section 6 could be replaced by the much simpler condition $G(t_1, t_2) = 0$ if $t_2 \geqq \sqrt{t_1}$. However, as remarked in section 6, even this property has not yet been proved.

Acknowledgements. The writer is highly indebted to Dr. Erich L. Lehmann, whose inspiring lectures laid the foundation for this study, and whose constructive criticism was always deeply appreciated. A special word of thanks is due Dr. Abraham Seidenberg for making available Theorem 1 in section 5. Gratitude also goes to Dr. Jerzy Neyman for continuous interest in this work. Finally, the writer wants to thank Dr. J. Kiefer and the referee for many valuable suggestions.

REFERENCES

[1] ALICE A. ASPIN, "An examination and further development of a formula arising in the problem of comparing two mean values," *Biometrika*, Vol. 35 (1948), pp. 88–96.

[2] R. R. BAHADUR, "Sufficiency and statistical decision functions," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 423–462.

[3] S. KARLIN, "Decision theory for Pólya type distributions. Case of two actions, I." *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability,*" Vol. 1, University of California Press, 1956, pp. 115–128.

[4] W. KRUSKAL, "The monotonicity of the ratio of two non-central t density functions," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 162–165.

[5] E. L. LEHMANN, "Some principles of the theory of testing hypotheses," *Ann. Math. Stat.*, Vol. 21 (1950), pp. 1–26.

[6] E. L. LEHMANN, "Theory of testing hypotheses," mimeographed notes recorded by Colin Blyth, ASUC Bookstore, University of California, Berkeley, 1949.

[7] E. L. LEHMANN, "Ordered families of distributions," *Ann. Math. Stat.*, Vol. 26 (1955), pp. 399–419.

[8] E. L. LEHMANN AND H. SCHEFFÉ, "Completeness, similar regions, and unbiased estimation. Part I," *Sankhyā*, Vol. 10 (1950), pp. 305–340.

[9] E. L. LEHMANN AND H. SCHEFFÉ, "Completeness, similar regions, and unbiased estimation. Part II," *Sankhyā*, Vol. 15 (1955), pp. 219–236.

[10] J. NEYMAN, "Outline of a theory of statistical estimation based on the classical theory of probability," *Phil. Trans. Roy. Soc., London, Series A*, Vol. 236 (1937), pp. 333–380.

[11] J. NEYMAN AND E. S. PEARSON, "On the problem of the most efficient tests of statistical hypotheses," *Phil. Trans. Roy. Soc., London, Series A*, Vol. 231 (1933), pp. 289–337.

[12] P. B. PATNAIK, "Hypotheses concerning the means of observations in normal samples," *Sankhyā*, Vol. 15 (1955), pp. 343–372.

[13] H. SCHEFFÉ, "On solutions of the Behrens-Fisher problem, based on the t-distribution," *Ann. Math. Stat.*, Vol. 14 (1943), pp. 35–44.

[14] A. WALD, "Testing the difference between the means of two normal populations with unknown standard deviations," *Selected Papers in Statistics and Probability*, Stanford University Press, 1957, pp. 669–695.

[15] B. L. WELCH, "On the studentization of several variances," *Ann. Math. Stat.*, Vol. 18 (1947), pp. 118–122.

[16] R. A. WIJSMAN, "Incomplete sufficient statistics and similar tests," (abstract), *Ann. Math. Stat.*, Vol. 27 (1956), p. 864.

[17] J. WOLFOWITZ, "The power of the classical tests associated with the normal distribution," *Ann. Math. Stat.*, Vol. 20 (1949), pp. 540–551.

Appendix 1. PROOF OF THEOREM 1 (Seidenberg). For the purpose of this proof we shall replace the s_i by x_i. Let $P_i = A_i(\theta)x_i^{d_i} + \cdots$. Let $d = \max\{d_i\}$. Multiplying P_i by $x_i^{d-d_i}$, we may suppose all the d_i equal. Multiplying P_1 by

$A_2 \cdot \cdots \cdot A_m$, P_2 by $A_1 A_3 \cdot \cdots \cdot A_m$, etc., we may suppose all the A_i equal. Now we have $P_i = \Delta(\theta)x_i^d + \cdots$, $i = 1, \cdots, m$, where Δ is some polynomial in $\theta_1, \cdots, \theta_k$.

Suppose we have a congruence of the form

$$\Pi x_i^{r_i} \Delta^{\Sigma r_i} \equiv R(x, \theta) \bmod (P_1, \cdots, P_m)$$

(i.e. the two sides are equal whenever all P_i vanish), with R a polynomial in the x's and θ's. Let $M = \max \{\deg_\theta P_i\}$. The left hand side has degree in the θ's at most $M\Sigma r_i$. Assume this to be the case also for $R(x, \theta)$. Assume further that $\deg_{x_i} R(x, \theta) \leqq d - 1$, $i = 1, \cdots, m$. Multiplying the congruence by $x_j\Delta$, on the left we get a power product of degree $1 + \Sigma r_i$ in the x_i times $\Delta^{1+\Sigma r_i}$. On the right there possibly appears a power x_j^d: if so, we replace Δx_j^d by

$$(\Delta x_j^d - P_j) \bmod P_j .$$

In this way we get a congruence

$$\Pi x_i^{s_i} \Delta^{\Sigma s_i} \equiv R'(x, \theta) \bmod (P_1, \cdots, P_m)$$

with $\Sigma s_i = 1 + \Sigma r_i$, $\deg_{x_i} R' \leqq d - 1$ $(i = 1, \cdots, m)$, $\deg_\theta R' \leqq M\Sigma s_i$. The congruences

$$x_i^d \Delta^d \equiv \Delta^{d-1}(x_i^d - P_i) \bmod (P_1, \cdots, P_m)$$

are of the above form. Multiplying by various power products of the $x_j\Delta$, we again get congruences of the stated form. Let $s \geqq s_0 = m(d - 1) + 1$. Then any power product of the x_i of degree s must have a factor x_i^d for at least one i. Hence we can get a congruence of the desired form with any power product of the $x_j\Delta$ of degree s on the left. For any such power product there may be several congruences: choose one.

For a fixed integer $\gamma \geqq s_0$ (to be determined in a moment), we consider all the power products of the $x_i\Delta$ of degree between s_0 and γ; and all the congruences, one for each power product. We still multiply each of these by an appropriate power of Δ so that Δ^γ is the power of Δ occurring on the left. On the right, then, all polynomials are of degree $\leqq M\gamma$ in the θ's and of degree $\leqq d - 1$ in each x_i.

Let $N(p, q)$ be the number of distinct power products of degree p or less in q letters. Then $N(p, q) = (p + q)(p + q - 1) \cdots (p + 1) / q!$ We are considering, then, $N(\gamma, m) - N(s_0 - 1, m)$ congruences. The right hand sides of these congruences are linear combinations over K of power products of degree $\leqq M\gamma$ in the θ's and of degree $\leqq m(d - 1)$ in the x's; therefore in at most $N(M\gamma, k)N(m(d - 1), m)$ power products. Since

$$\deg_\gamma [N(\gamma, m) - N(s_0 - 1, m)] = m > k = \deg_\gamma N(M\gamma, k)N(m(d - 1), m),$$

we see that for sufficiently large γ,

$$N(\gamma, m) - N(s_0 - 1, m) > N(M\gamma, k)N(m(d - 1), m).$$

Let γ be taken large enough for this to be realized. Then there exist $c_{i_1,\cdots,i_m} \varepsilon K$, not all $= 0$, such that

$$\Delta^\gamma \Sigma c_{i_1,\cdots,i_m} x_1^{i_1} \cdot \cdots \cdot x_m^{i_m} \equiv 0 \bmod (P_1, \cdots, P_m). \qquad \text{Q.E.D.}$$

Appendix 2. It will be proved here that the integrated terms (26)–(28) vanish in the limit $B \to \infty$, then $A \to \infty$. Since ϕ is similar and of size α, the function φ defined by (22) has the property

$$\int_0^\infty \int_{-\sqrt{t_1}}^{\sqrt{t_1}} \varphi(t_1, t_2) \exp\left[-\frac{1}{2\sigma^2} t_1 + \frac{r_0}{\sigma} t_2\right] dt_2\, dt_1 \underset{\sigma}{\equiv} 0. \tag{32}$$

This property is crucial for showing (26) $\to 0$, but is not needed for (27) and (28).

We shall first treat (26). Since σ is an arbitrary positive number, we shall give the proof with σ replaced by $r\sigma/r_0$, which will be useful for later purposes. With this change we substitute (23) into (26) and get

$$\begin{aligned}
&\int_{-\infty}^{B} G(A, t_2) \exp\left[-\frac{r_0^2}{r^2}\frac{A}{2\sigma^2} + \frac{r_0}{\sigma} t_2\right] dt_2 \\
&\quad = \exp\left[-\frac{r_0^2}{r^2}\frac{A}{2\sigma^2}\right] \int_0^A dt_1' \int_{-\sqrt{t_1'}}^{\sqrt{t_1'}} \varphi(t_1', t_2')\, dt_2' \int_{-\infty}^{B} (A - t_1')^{-1/2} \\
&\qquad\qquad \cdot \exp\left[-\frac{r_0^2}{2}\frac{(t_2 - t_2')^2}{A - t_1'} + \frac{r_0}{\sigma} t_2\right] dt_2 \\
&\quad = \frac{\sqrt{2\pi}}{r_0} \exp\left[\left(1 - \frac{r_0^2}{r^2}\right)\frac{A}{2\sigma^2}\right] \int_0^A dt_1' \int_{-\sqrt{t_1'}}^{\sqrt{t_1'}} \varphi(t_1', t_2') \\
&\qquad \cdot \exp\left[-\frac{1}{2\sigma^2} t_1' + \frac{r_0}{\sigma} t_2'\right] dt_2' \int_{-\infty}^{B} \frac{r_0}{\sqrt{2\pi}} (A - t_1')^{-1/2} \\
&\qquad\qquad \cdot \exp\left[-\frac{r_0^2}{2}\frac{(t_2 - (A - t_1')/\sigma r_0 - t_2')^2}{A - t_1'}\right] dt_2
\end{aligned}$$

The integral over t_2 can be written

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{B'} \exp\left[-\frac{1}{2} z^2\right] dz,$$

in which

$$B' = r_0(A - t_1')^{-1/2}(B - (A - t_1')/\sigma r_0 - t_2').$$

As $B \to \infty$, $B' \to \infty$ and the integral converges monotonically increasing to 1. The integration over t_2' and t_1' can be considered as a double integral of the form $\iint f_B(t_1', t_2')\, dt_1'\, dt_2'$, in which f_B is bounded in absolute value by

$$|\varphi(t_1', t_2')| \exp\left[-\frac{1}{2\sigma^2} t_1' + \frac{r_0}{\sigma} t_2'\right]$$

which is integrable. Applying the Lebesgue bounded (dominated) convergence theorem, we may take the limit as $B \to \infty$ under the integral. We have

$$\lim_{B\to\infty} f_B(t_1', t_2') = \int_0^A dt_1' \int_{-\sqrt{t_1'}}^{\sqrt{t_1'}} \varphi(t_1', t_2') \exp\left[-\frac{1}{2\sigma^2}t_1' + \frac{r_0}{\sigma}t_2'\right] dt_2'.$$

By (32) we may replace the integral on the right in the above equation by minus the integral with same integrand but the t_1' integration running from A to ∞. Thus we get

$$\lim_{B\to\infty} \int_{-\infty}^{B} G(A, t_2) \exp\left[-\frac{r_0^2}{r^2}\frac{A}{2\sigma^2} + \frac{r_0}{\sigma}t_2\right] dt_2 = \xi(A)$$

with

$$\xi(A) = -\frac{\sqrt{2\pi}}{r_0} \exp\left[\left(1 - \frac{r_0^2}{r^2}\right)\frac{A}{2\sigma^2}\right] \int_A^\infty dt_1' \int_{-\sqrt{t_1'}}^{\sqrt{t_1'}} \varphi(t_1', t_2')$$
$$\cdot \exp\left[-\frac{1}{2\sigma^2}t_1' + \frac{r_0}{\sigma}t_2'\right] dt_2'.$$

Since ϕ is bounded, we see by (22) that $\varphi(t_1', t_2')$ is bounded in absolute value by const. $h(t_1', t_2')$, which is bounded by const. $t_1'^{(n/2)-1}$. In the integration over t_2' we have that

$$\int_{-\sqrt{t_1'}}^{\sqrt{t_1'}} \exp\left[\frac{r_0}{\sigma}t_2'\right] dt_2'$$

is bounded by $2\sqrt{t_1'} \exp\,[(r_0/\sigma)\sqrt{t_1'}]$. Thus

$$|\,\xi(A)\,| < \text{const.} \exp\left[\left(1 - \frac{r_0^2}{r^2}\right)\frac{A}{2\sigma^2}\right] \int_A^\infty t_1'^{(n-1)/2} \exp\left[-\frac{1}{2\sigma^2}t_1' + \frac{r_0}{\sigma}\sqrt{t_1'}\right] dt_1'.$$

We make the substitutions $t_1' = \sigma^2(u + r_0)^2$, $A = \sigma^2K^2$, then A and K go to ∞ together. Put $\xi(A) = \eta(K)$, then

$$|\,\eta(K)\,| < \text{const.} \exp\left[\left(1 - \frac{r_0^2}{r^2}\right)\frac{K^2}{2}\right] \int_{K-r_0}^\infty (u + r_0)^n \exp\,[-\tfrac{1}{2}u^2]\, du.$$

In the integrand, $(u + r_0)^n$ can be bounded by const. u^n, and by partial integration one finds that

$$\int_{K-r_0}^\infty u^n \exp\,[-\tfrac{1}{2}u^2]\, du$$

is bounded by const. $K^{n-1} \exp\,[-\frac{1}{2}(K - r_0)^2]$. This leads then to

$$|\,\eta(K)\,| < \text{const.}\, K^{n-1} \exp\left[-\frac{r_0^2}{r^2}\frac{K^2}{2} + r_0K\right]$$

which $\to 0$ as $K \to \infty$. Q.E.D.

Of the integrated terms (27) and (28) we shall only treat (28), since (27) is a little simpler and follows the same pattern. It can be shown that (23) can be differentiated partially with respect to t_2 under the integral sign, provided $t_2^2 > t_1$. Substituting the result into (28) we obtain, apart from a multiplicative constant,

$$\iiint (t_1 - t_1')^{-3/2}(B - t_2')\varphi(t_1', t_2') \cdot \exp\left[-\frac{1}{2\sigma^2} t_1 + \frac{r}{\sigma} B - \frac{r_0^2}{2} \frac{(B - t_2')^2}{t_1 - t_1'}\right] dt_2'\, dt_1'\, dt_1 \tag{33}$$

in which the integration is over the region $t_2'^2 \leqq t_1' \leqq t_1 \leqq A$. We shall show that (33) $\rightarrow 0$ as $B \rightarrow \infty$ for fixed A, after which taking the limit $A \rightarrow \infty$ yields then trivially 0. Clearly the integrand in (33) approaches 0 as $B \rightarrow \infty$. It suffices therefore to show that limit and integral may be interchanged. By the Lebesgue bounded convergence theorem it is sufficient to show that the integrand is bounded in absolute value by an integrable function independent of B (but possibly dependent on A). Let $B_0 > \sqrt{A}$ and consider only values of $B \geqq B_0$. The integrand is bounded in absolute value by $|\varphi(t_1', t_2')| f_1 f_2 f_3$, in which

$$f_1 = \exp\left[\frac{r}{\sigma} B - \frac{r_0^2}{4} \frac{(B - t_2')^2}{t_1 - t_1'}\right], \qquad f_2 = \exp\left[-\frac{r_0^2}{4} \frac{(B - t_2')^2}{t_1 - t_1'}\right]\left(\frac{B - t_2'}{\sqrt{t_1 - t_1'}}\right)^3,$$

and $f_3 = (B - t_2')^{-2}$. Now f_1 is bounded by

$$\exp\left[\frac{r}{\sigma} B - \frac{r_0^2}{4} \frac{(B - \sqrt{A})^2}{A}\right]$$

which is bounded by a constant; f_2 is of the form $y^3 \exp[-(r_0^2/4)y^2]$ and is therefore also bounded by a constant; f_3 is bounded by the constant $(B_0 - \sqrt{A})^{-2}$. Finally we have then that the integrand in (33) is bounded in absolute value by const. $|\varphi(t_1', t_2')|$, which is integrable over the bounded region $t_2'^2 \leqq t_1' \leqq t_1 \leqq A$. Q.E.D.

Appendix 3. We will show that if ϕ is invariant, then $G = 0$ in the region $t_2 \geqq \sqrt{t_1}$. Let $y = t_2/\sqrt{t_1}$, and $y' = t_2'/\sqrt{t_1'}$. If ϕ is invariant, it is a function of y only. Put $\phi(t_1, t_2) = \phi^*(y)$, so that by (22) and (18)

$$\varphi(t_1, t_2) = \text{const. } t_1^{(n/2)-1}(1 - y^2)^{(n/2)-1}(\phi^*(y) - \alpha).$$

After substitution into (23) and making the change of variable $\tau = t_1'/t_1$, we can write (23) as

$$G(t_1, t_2) = \text{const. } t_1^{n/2} \exp\left[-\frac{r_0^2}{2} y^2\right] \cdot \int_0^1 (1 - y'^2)^{(n/2)-1}(\phi^*(y') - \alpha) f(y, y')\, dy', \tag{34}$$

in which

$$f(y, y') = \int_0^1 \tau^{(n-1)/2}(1 - \tau)^{-1/2} \exp\left[-\frac{r_0^2}{2} \frac{y^2\tau - 2yy'\sqrt{\tau} + y'^2\tau}{1 - \tau}\right] d\tau \tag{35}$$

Throughout we restrict y and y' to $y > 1$, $y' \leqq 1$. Let the differential operator D_y be defined as

$$D_y = \frac{1}{r_0^2}\frac{\partial^2}{\partial y^2} - y\frac{\partial}{\partial y} - (n+1) \tag{36}$$

and the operator $D_{y'}$ similarly by replacing in (36) y by y'. Then f satisfies the two equations

$$D_y f(y, y') = 0 \tag{37}$$

$$D_{y'} f(y, y') = 0. \tag{38}$$

Furthermore, it can be seen from (35) that $f \to 0$ if $y \to \infty$ for fixed y', or if $y' \to -\infty$ for fixed y. Two linearly independent solutions of the equation

$$D_y u(y) = 0 \tag{39}$$

are u_1 and u_2, with $u_2(y) = u_1(-y)$, and

$$u_1(y) = \int_0^\infty t^{(n-1)/2} \exp\left[-\tfrac{1}{2}t + r_0\sqrt{t}\, y\right] dt \tag{40}$$

When $y \to \infty$, $u_1(y) \to \infty$ whereas $u_2(y) \to 0$, with the opposite behavior as $y \to -\infty$. It follows from (37) and (38), from the behavior of the functions u_1 and u_2, and from the behavior of f as $y \to \infty$ or $y' \to -\infty$, that f must equal

$$f(y, y') = \text{const. } u_1(y')u_2(y). \tag{41}$$

Substituting (41) into (34), it remains to be shown that the integral

$$\int_0^1 (1 - y'^2)^{(n/2)-1}(\phi^*(y') - \alpha)u_1(y')\, dy' \tag{42}$$

equals 0, with u_1 given by (40). Replacing in (40) t by t_1 and in (42) y' by $t_2/\sqrt{t_1}$, the integral (42) is nothing else but the expectation of $\phi - \alpha$ with respect to the distribution specified by $r = r_0$, $\sigma = 1$. Since ϕ is similar of size α this expectation vanishes, Q.E.D.

A METHOD OF GENERATING BEST ASYMPTOTICALLY NORMAL ESTIMATES WITH APPLICATION TO THE ESTIMATION OF BACTERIAL DENSITIES[1]

By Thomas S. Ferguson[2]

University of California, Berkeley

0. Summary. Various minimum χ^2 methods used for generating B.A.N. estimates are summarized, and a new method which generates B.A.N. estimates as roots of certain linear forms is introduced and investigated. As a particular application of the method, the estimation of the bacterial density in an experiment using dilution series is considered.

1. Introduction. The purpose of the present paper is to describe a simple method by which estimates having the usual asymptotic properties of Best Asymptotically Normal (B.A.N.) estimates can be obtained.

Originally B.A.N. estimates were introduced by J. Neyman [1] for a situation in which the underlying probability distributions have a multinomial-like structure. This was followed by a paper by E. W. Barankin and J. Gurland [2] who extended the class of estimation problems for which B.A.N. estimates could be used and also described quite general methods of generation of such estimates. Other results in this direction have been obtained by C. L. Chiang [3] and L. Le Cam [4] and W. Taylor [5].

A best asymptotically normal estimate θ^* of a parameter θ is, loosely speaking, one which is asymptotically normally distributed about the true parameter value, and which is best in the sense that out of all such asymptotically normal estimates it has the least possible asymptotic variance. Thus a B.A.N. estimate will be asymptotically the "most accurate" estimate of a parameter; but the value to a statistician of obtaining such estimates is even greater than is indicated by this. In the aforementioned paper of Neyman, a simple method of testing hypotheses is described which is asymptotically equivalent to the likelihood ratio test and involves the use of the χ^2 statistic and a B.A.N. estimate. It usually turns out that the hardest work in applying this technique is in computing the estimate. Thus it is important to have a number of different methods for computing B.A.N. estimates available to the applied statistician. The usual methods of obtaining B.A.N. estimates will be summarized briefly in section 2.

The objective of all these methods is at least in part a practical one and is essentially two-fold. First, it is hoped that some of these estimates will be easily computable. Second, even though all these estimates have the same

---

Received July 11, 1956; revised May 7, 1958.

[1] This paper was prepared with the support of the Office of Ordnance Research, U. S. Army under Contract DA-04-200-ORD-171.

[2] Now at the University of California at Los Angeles.

asymptotic properties, they may differ widely in their small sample properties, and it seems reasonable that the choice of the proper estimate to use should depend in part on the behavior of the estimate for small samples. As a consequence, a large class of estimates with best asymptotic properties is proposed with the hope that some of the easily computable estimates will have small sample properties which are reasonably good. Blind adherence to the principle of maximum likelihood, for example, may lead to more difficult computations and still yield less accurate estimates than other methods of estimation.

A new approach to generating B.A.N. estimates as roots of linear forms of certain variables is suggested in section 3. In cases where minimum distance methods are applicable, the procedure proposed here leads to estimates which are solutions of equations obtained by simplifying in a suitable manner the equations obtained by the original methods. By way of an example, section 4 contains an application of this approach to the problem of estimation of bacterial density.

2. A review of the minimum χ^2 methods of generating B.A.N. estimates. Since the following methods are to be found in the literature at various levels of generalities, a complete mathematical description of the hypotheses necessary for their validity will be omitted.

Let $X_1, X_2, \cdots, X_n, \cdots$ be a sequence of independent identically distributed s-dimensional random vectors whose distribution depends upon a parameter θ belonging to an open subset Θ of k-dimensional Euclidean space R_k with $k \leqq s$. Let $P(\theta) = E(X \mid \theta)$ be the s-dimensional vector of the expectations of the vector X_n, and let $\Sigma(\theta) = var(X \mid \theta) = E\{[X - P(\theta)]\,[X - P(\theta)]'\}$ be the $s \times s$ covariance matrix which is assumed to be finite and non-singular for each $\theta \varepsilon \Theta$. Furthermore, it is assumed that $P(\theta)$ is a one-to-one bicontinuous map from Θ to a subset of s-dimensional Euclidean space with continuous partial derivatives of the second order. Let Z_n be the s-dimensional random vector defined by $nZ_n = \sum_{j=1}^{n} X_j$.

The quadratic form

$$n[Z_n - P(\theta)]'\,\Sigma(\theta)^{-1}\,[Z_n - P(\theta)] \tag{2.1}$$

will be designated by the name of χ^2. The value $\mathring{\theta}(Z_n)$ of θ which minimizes this quadratic form will be called the minimum χ^2 estimate of θ. An as example take the multinomial case where there are n independent trials each capable of producing any of $s + 1$ possible outcomes. Let the probability on each successive trial be $p_i(\theta)$ of producing the ith outcome. Let z_i denote the proportion of the trials which result in the ith outcome. Then

$$\chi^2 = n \sum_{i=1}^{s+1} \frac{(z_i - p_i(\theta))^2}{p_i(\theta)} \tag{2.2}$$

is the familiar Pearson χ^2. It may be shown that (2.2) is algebraically equal to the χ^2 of the form (2.1) where the vector Z_n is the vector of the first s z_i's. The ad-

vantage of (2.1) lies in the fact that it describes a method for estimating parameters of continuous distributions.

Barankin and Gurland [2] have shown that the minimum χ^2 estimate, as defined above, is B.A.N. where the X_n have distributions belonging to a Koopman's family, and Z_n is a vector of sufficient statistics. When the distributions under consideration do not form a Koopman's family with sufficient statistics Z_n, the term B.A.N. estimate is perhaps not entirely justifiable but will be retained for convenience. The precise definition of B.A.N. estimate to be adopted is somewhat irrelevant, because the methods reviewed in this section and the method developed in section 3, give estimates which have the same asymptotic behavior as the minimum χ^2 estimates. In section 3.3, the sense in which the estimates are best is stated more precisely.

Starting with this basic minimum χ^2 estimate, several methods may be used to generate large classes of estimates. These methods will be described below. Method I is due essentially to Karl Pearson. Method II as a general method may be found in Barankin and Gurland [2] and Taylor [5], but special cases were used earlier (see Berkson [6]). Method III evolved from practical work and is of unknown authorship. Method IV is due to Neyman [1].

Method I. Modification. Let $M_n(Z_n, \theta)$ be an $s \times s$ symmetric positive definite matrix. The quadratic form

$$Q_n(\theta) = n[Z_n - P(\theta)]'M_n(Z_n, \theta)[Z_n - P(\theta)] \tag{2.3}$$

will be called the modified or reduced χ^2. The estimate $\hat{\theta}_M(Z_n)$ which minimizes the modified χ^2 with the function $M_n(Z_n, \theta)$ depending only on Z_n and not on θ or n, will be called the minimum modified χ^2 estimate of θ. For example, the estimate which minimizes the Pearson modified χ^2,

$$\chi_M^2 = n \sum_{i=1}^{s+1} \frac{(z_i - P_i(\theta))^2}{z_i} \tag{2.4}$$

is such an estimate.

Under the condition that $M_n(Z_n, \theta) \to \Sigma^{-1}(\theta)$ in probability as $n \to \infty$ when θ is the true value of the parameter, and under certain regularity conditions, the minimum modified χ^2 estimate of θ will have the same asymptotic properties as the minimum χ^2 estimate of θ (or simply $\theta_M(Z_n)$ will be B.A.N., according to the conventions made.)

Method II. Transformation. Let $g(x)$ be any function from R_s to R_s with continuous first partial derivatives

$$g(x) = \begin{pmatrix} g_1(x_1, \cdots, x_s) \\ \vdots \\ g_s(x_1, \cdots, x_s) \end{pmatrix} \tag{2.5}$$

Let the $s \times s$ matrix of first partial derivatives be denoted by

$$\dot{g}(x) = \begin{pmatrix} \frac{\partial}{\partial x_1} g_1 & \cdots & \frac{\partial}{\partial x_1} g_s \\ \vdots & & \\ \frac{\partial}{\partial x_s} g_1 & \cdots & \frac{\partial}{\partial x_s} g_s \end{pmatrix} \tag{2.6}$$

We shall call the quadratic form

$$n[g(Z_n) - g(P(\theta))]'[\dot{g}(P(\theta))\Sigma(\theta)\dot{g}(P(\theta))']^{-1}[g(Z_n) - g(P(\theta))] \tag{2.7}$$

the transformed χ^2. More generally, we may consider the combinations of Methods I and II, and replace the matrix of the quadratic form (2.7) by an estimate

$$Q_n(\theta) = n[g(Z_n) - g(P(\theta))]'M_n(Z_n, \theta)[g(Z_n) - g(P(\theta))]. \tag{2.8}$$

We assume that $M_n(Z_n, \theta) \to [\dot{g}(P(\theta))\Sigma(\theta)\dot{g}(P(\theta))']^{-1}$ in probability and the regularity conditions needed for Method I. In addition, one needs regularity conditions on g, namely that g is a one-to-one bicontinuous map from a neighborhood of $P(\Theta)$ into R_s, with continuous partial derivatives of the second order and that the matrix $\dot{g}(P(\theta))$ is nonsingular for each $\theta \varepsilon \Theta$. Then the minimum transformed χ^2 estimates, that is the value $\hat{\theta}_T(Z_n)$ of θ minimizing (2.7), will be a B.A.N. estimate of θ.

This method of generating B.A.N. estimates also applies to the χ^2 of (2.2); for example, letting $g_i(x)$ be the real-valued transformation applied to the ith cell

$$\chi^2 = n \sum_{i=1}^{s+1} \frac{[g_i(z_i) - g_i(p_i(\theta))]^2}{p_i(\theta)g_i'(p_i(\theta))} \tag{2.9}$$

or modified,

$$\chi^2 = n \sum_{i=1}^{s+1} \frac{[g_i(z_i) - g_i(p_i(\theta))]^2}{z_i\, g_i'(z_i)^2}. \tag{2.10}$$

The well-known example of Berkson [6] is of the type (2.10).

Many times the functions g_i may be chosen so that $g_i(p_i(\theta))$ is a linear function of the parameters $\theta_1, \cdots, \theta_k$. In such cases finding the value of θ which minimizes the χ^2 of equation (2.10) results in solving k linear equations in k unknowns. The reader may consult the paper of W. Taylor [5] for examples.

Method III. Expansion in a Taylor series about a $O(\sqrt{n})$-consistent estimate. An estimate θ_n^* of θ will be called $O(\sqrt{n})$-consistent if $\sqrt{n}(\theta_n^* - \theta)$ is bounded in probability uniformly in n when θ is the true value of the parameter; that is, for every $\epsilon > 0$ and $\theta \varepsilon \Theta$, there exists a number B so large that for every $n = 1, 2, \cdots$

$$P[\sqrt{n} \mid \theta_n^* - \theta \mid > B \mid \theta] < \epsilon. \tag{2.11}$$

Many types of estimates satisfy this requirement. For example, under certain regularity conditions, estimation by the method of moments yields estimates θ_n^* for which $\sqrt{n}(\theta_n^* - \theta)$ is asymptotically normal when θ is the true value of the parameter. This follows from a theorem of Cramér [7], p. 366, which states that certain functions of the moments are asymptotically normal. Such asymptotically normal estimates as this are obviously $O(\sqrt{n})$-consistent.

One may try to apply a correction to θ_n^* by an application of the method of expansion in a Taylor series to get an estimate closer to the minimum χ^2 estimate. It is known, however, that one such application to a $O(\sqrt{n})$-consistent estimate will give a B.A.N. estimate. More specifically, consider the expansion of some one of the previously mentioned χ^2's (modified and/or transformed) in a Taylor series to the second degree terms about a $O(\sqrt{n}$-consistent estimate θ_n^* of θ.

$$\chi^2(\theta) = \chi^2(\theta_n^*) + \dot{\chi}^2(\theta_n^*)'(\theta - \theta_n) + \tfrac{1}{2}(\theta - \theta_n^*)'\ddot{\chi}^2(\theta_n^*)(\theta - \theta_n^*) + \text{Rem.} \tag{2.12}$$

where $\dot{\chi}^2(\theta)$ is the $k \times 1$ vector of first derivatives of $\chi^2(\theta)$ and $\ddot{\chi}^2(\theta)$ is the $k \times k$ matrix of second derivatives of $\chi^2(\theta)$.

$$\dot{\chi}^2(\theta) = \begin{pmatrix} \dfrac{\partial}{\partial\theta_1}\chi^2(\theta) \\ \vdots \\ \dfrac{\partial}{\partial\theta_k}\chi^2(\theta) \end{pmatrix} \tag{2.13}$$

$$\ddot{\chi}^2(\theta) = \begin{pmatrix} \dfrac{\partial^2}{\partial\theta_1^2}\chi^2(\theta) & \cdots & \dfrac{\partial^2}{\partial\theta_1\,\partial\theta_k}\chi^2(\theta) \\ \vdots & & \\ \dfrac{\partial^2}{\partial\theta_k\,\partial\theta_1}\chi^2(\theta) & \cdots & \dfrac{\partial^2}{\partial\theta_k^2}\chi^2(\theta) \end{pmatrix}. \tag{2.14}$$

Instead of finding that value of θ which minimizes $\chi^2(\theta)$, one may discard the remainder term and find that value $\hat{\theta}_n$ of θ which minimizes the first three terms of the expansion. This estimate $\hat{\theta}_n$ will then be a B.A.N. estimate of θ. This method of generating B.A.N. estimates is important because it leads to k linear equations in k unknowns and is thus comparatively easy to apply.

Method IV. Linearization of the side conditions. This method, due to Neyman [1], was proposed with the specific intention of finding a B.A.N. estimate which could be computed by solving linear equations. In minimizing some χ^2 like (2.1), one may consider the vector P as the vector of parameters which are subject to certain restrictions, called side conditions, due to the dependence of P on θ. If there are s independent components of the vector P and k parameters, there will be $s - k$ side conditions on the p's.

$$f_j(p_1, \cdots, p_s) = 0 \qquad \text{for} \quad j = 1, \cdots, k - s \tag{2.16}$$

One may then minimize χ^2 subject to these side conditions by the method of Lagrange multipliers. However, a simpler procedure would be to minimize χ^2 subject to the linearized counterpart of (2.15), that is, the first two terms of the Taylor series expansion about the point z_n. The solution for the estimate then only requires solution of linear equations. For a fuller account of the subject, the reader should consult the papers of Neyman and of Barankin and Gurland. The outline of the method given here is added only for the sake of completeness and no mention of the method will be made in the later sections of the paper.

3. B.A.N. estimates as roots of linear forms. The method customarily used to find a minimum χ^2 estimate is to differentiate χ^2 with respect to each of the parameters separately, set the results equal to zero and solve the resulting system of simultaneous equations. For example, one may differentiate the χ^2 of the equation (2.4) and obtain the equations

$$-2n \sum_{i=1}^{s} \frac{z_i - p_i(\theta))}{z_i} \frac{\partial p_i(\theta)}{\partial \theta_j} = 0 \qquad \text{for } j = 1, 2, \cdots, k, \tag{3.1}$$

or one may differentiate the χ^2 of equation (2.3) with $M_n(Z_n, \theta)$ a function of Z_n only, such that $M_n(Z_n) \rightarrow \Gamma(\theta)$ in probability and the regularity conditions hold, and obtain

$$-n\dot{P}(\theta)M(Z_n)(Z_n - P(\theta)) = 0 \tag{3.2}$$

where $\dot{P}(\theta)$ is the $k \times s$ matrix of first partial derivatives of the vector $P(\theta)$,

$$\dot{P}(\theta) = \begin{pmatrix} \dfrac{\partial P_1(\theta)}{\partial \theta_1} & \cdots & \dfrac{\partial P_s(\theta)}{\partial \theta_1} \\ \vdots & & \\ \dfrac{\partial P_1(\theta)}{\partial \theta_k} & \cdots & \dfrac{\partial P_s(\theta)}{\partial \theta_k} \end{pmatrix} \tag{3.3}$$

and the 0 is the $k \times 1$ vector with a zero term in each component so that (3.2) represents k equations in k unknowns.

Well-chosen roots to equations such as (3.1) and (3.2) are B.A.N. estimates of the unknown parameters. This suggests that instead of starting with a quadratic form in $(Z_n - P(\theta))$ and finding values of θ which make the form a minimum, it may be simpler to start with an arbitrary linear form in $(Z_n - P(\theta))$ and find the roots. Roots of certain such linear forms, namely, (3.1) and (3.2), will be B.A.N. estimates. Furthermore, such a method of generating B.A.N. estimates will probably satisfy the requirement that they be easy to compute. It is the purpose of this section to investigate the asymptotic distribution of roots of linear forms in $(Z_n - P(\theta))$, and the conditions for such roots to be B.A.N. estimates of the parameters.

3.1. Preliminary lemma. This section contains an implicit function theorem needed for the proof of the main theorem. First an implicit function theorem

which can be found in Pierpont [8], p. 293, for example is stated, from which the lemma of this section will follow. The unicity of the implicit function is stated in a somewhat stronger form than found in Pierpont. This strengthening can be obtained by modifying his proof slightly and the details of the proof need not be given here.

Let $F(x, u)$ be a function of variables $x \varepsilon R_s$ and $u \varepsilon R_k$ with values in R_k. Let $a \varepsilon R_s$ and $b \varepsilon R_k$, and assume that

(i) *$F(x, u)$ is continuous and $F_u(x, u)$ exists and is continuous in a neighborhood of the point (a, b).*

(ii) *$F(a, b) = 0$ and $F_u(a, b)$ is nonsingular. Then, there exists a neighborhood N of a, and a function $\phi(x)$ from R_s to R_k, such that*

(1) *$\phi(x)$ is continuous in N,*

(2) *$\phi(a) = b$,*

(3) *$F(x, \phi(x)) = 0$ for $x \varepsilon N$, and*

(4) *(uniqueness) there exists a neighborhood N' of the point b such that for $u \varepsilon N'$ and $x \varepsilon N$, $F(x, a) \neq 0$ unless $u = \phi(x)$.*

In the above theorem $F_u(x, u)$ represents the $k \times k$ matrix of partial derivatives of $F(x, u)$ with respect to u, as in equation (3.3). The assumption of continuity of $F_u(x, u)$ means that each component of the matrix is assumed to be continuous.

The following lemma is an extension of this theorem, similar to that found in Graves [9], p. 144, to the situation in which $F(x, u)$ is known to vanish along some curve in R_{s+k}, rather than just at one point.

LEMMA. *Let $F(x, u)$ be a function of variables $x \varepsilon R_s$ and $u \varepsilon R_k$ with values in R_k, $k \leqq s$. Let $p(u)$ be a function from some set $D \subset R_k$ to R_s, and assume that*

(i) *D is an open set,*

(ii) *$p(u)$ is one-to-one and inversely continuous from D into R_s,*

(iii) *there is a neighborhood of the curve $\{(p(u), u) : u \varepsilon D\}$ in which $F(x, u)$ is continuous and $F_u(x, u)$ exists and is continuous.*

(iv) *$F(p(u), u) = 0$ and $F_u(p(u), u)$ is nonsingular for every $u \varepsilon D$.*

Then, there exists a neighborhood N of the set $S = \{p(u) : u \varepsilon D\}$ and a function $\phi(x)$ from R_s to R_k such that

(a) *$\phi(x)$ is continuous in N,*

(b) *$\phi(p(u)) = u$ for $u \varepsilon D$,*

(c) *$F(x, \phi(x)) = 0$ for $x \varepsilon N$,* and

(d) *there exists a neighborhood of the curve $\{(p(u), u) : u \varepsilon D\}$ in which the only zeros of the function $F(x, u)$ are the points $(x, \phi(x))$.*

PROOF. From the previous implicit function theorem, for every $u \varepsilon D$, there is a neighborhood $N_{p(u)}$ of the point $p(u)$ and a function $\phi_u(x)$ from R_s to R_k such that

(1) $\phi_u(x)$ is continuous in $N_{p(u)}$,

(2) $\phi_u(p(u)) = u$,

(3) $F(x, \phi_u(x)) = 0$ for $x \varepsilon N_{p(u)}$, and

(4) for y in some neighborhood N_u of the point u, and $x \varepsilon N_{p(u)}$

$$F(x, y) \neq 0 \quad \text{unless} \quad y = \phi_u(x).$$

Using the inverse continuity of the function $p(u)$, and the continuity of the function $\phi_u(x)$, we may replace the neighborhoods $N_p(u)$ by spherical neighborhoods $N'_{p(u)}$ with the two additional properties that

(5) if $p(u_1) \varepsilon N'_{p(u_2)}$ for some u_1 and $u_2 \varepsilon D$, then $u_1 \varepsilon N_{u_2}$
and

(6) if $x \varepsilon N'_{p(u)}$ for some $u \varepsilon D$, then $\phi_u(x) \varepsilon N_u$.

Now consider spherical neighborhoods $N''_{p(u)}$ with radii equal to $\frac{1}{3}$ that of $N'_{p(u)}$, and let N denote $\bigcup_{u \varepsilon D} N''_{p(u)}$. The set N is then obviously a neighborhood of the set S.

We will show that if $x_0 \varepsilon N''_{p(u_1)} \cap N''_{p(u_2)}$, then $\phi_{u_1}(x_0) = \phi_{u_2}(x_0)$. For since $N''_{p(u_1)} \cap N''_{p(u_2)}$ is not empty, either $p(u_1) \varepsilon N'_{p(u_2)}$ or $p(u_2) \varepsilon N'_{p(u_1)}$. Suppose without loss of generality that the former is true; then since $u_1 \varepsilon N_{u_2}$ and

$$F(p(u_1), \phi_{u_2}(p(u_1))) = 0,$$

we have $\phi_{u_2}(p(u_1)) = u_1$. Furthermore, for $x \varepsilon N''_{p(u_1)} \cap N'_{p(u_2)}$, $\phi_{u_2}(x)$ is continuous and satisfies $F(x, \phi_{u_2}(x)) = 0$; but $\phi_{u_1}(x) \varepsilon N_{u_1}$ for $x \varepsilon N'_{p(u_1)}$ and thus $\phi_{u_1}(x)$ is the unique function, continuous in $N''_{p(u_1)}$ and such that

$$\phi_{u_1}(p(u_1)) = u_1 \qquad \text{and} \qquad F(x, \phi_{u_1}(x)) = 0.$$

Hence, $\phi_{u_1}(x_0) = \phi_{u_2}(x_0)$.

Thus for $x \varepsilon N$ we may define $\phi(x) = \phi_u(x)$ for any u for which $x \varepsilon N''_{p(u)}$, since such a definition is unique. Now parts (a), (b), and (c) of the conclusion of the lemma are obvious. As for (d), the neighborhood can be chosen to be $\bigcup_{u \varepsilon D}[N''_{p(u)} \times N_u]$.

3.2. The main theorem. Let Z_n, $n = 1, 2, \cdots$ be a sequence of s-dimensional random vectors whose distribution depends upon a parameter θ in some set $\Theta \subset R_k$, $k \leqq s$. Let $P(\theta)$ be a function from Θ to R_s.

ASSUMPTION 1. *Θ is an open set.*

ASSUMPTION 2. *$\mathfrak{L}\{\sqrt{n}(Z_n - P(\theta)) \mid \theta\} \to \mathfrak{L}(Z)$ where Z is a normal random vector with mean zero and variance-covariance matrix $\Sigma(\theta)$. (That is,*

$$EZ = O, \quad EZZ' = \Sigma(\theta).)$$

The convergence used above is convergence in law or in distribution. Assumption 2 states that when θ is the true value of the parameter, the distribution of $\sqrt{n}(Z_n - P(\theta))$ converges to a normal distribution with mean zero and variance-covariance matrix $\Sigma(\theta)$. The law degenerate at some point a will be denoted by $\mathfrak{L}(a)$. Thus $\mathfrak{L}(X_n) \to \mathfrak{L}(a)$ means that X_n converges in probability to a.

ASSUMPTION 3. *The mapping $P(\theta)$ from Θ into R_s is homeomorphic (that is, one-to-one and bicontinuous) and continuously differentiable.*

Let $f(x, \theta)$ be a $k \times s$ matrix for each $x \varepsilon R_s$ and $\theta \varepsilon \Theta$.

ASSUMPTION 4. *There is a neighborhood $N_0 \subset R_s \times \Theta$ of the set*

$$\{(P(\theta), \theta) : \theta \varepsilon \Theta\}$$

within which $f(x, \theta)$ *and* $\partial/\partial\theta_j f(x, \theta)$ *for* $j = 1, 2, \cdots, k$ *are continuous jointly in* (x, θ).

Let $b(\theta) = f(P(\theta), \theta)$ and let $\dot{P}(\theta)$ be the $k \times s$ matrix of partial derivatives of $P(\theta)$, given by equation (3.3).

ASSUMPTION 5. *The matrix* $\dot{P}(\theta)b(\theta)'$ *is nonsingular for each* $\theta \varepsilon \Theta$.

Let

$$F(x, \theta) = f(x, \theta)(x - P(\theta)). \tag{3.4}$$

This is the linear form which will be used in the sequel to generate B.A.N. estimates of the parameter θ. The following theorem shows immediately that the root to the equation $F(Z_n, \theta) = 0$ will be a $O(\sqrt{n})$-consistent estimate of θ.

THEOREM 1. *Under assumptions* 1 *through* 5, *there exists a neighborhood* N *of the set* $S = \{P(\theta):\theta \varepsilon \Theta\}$ *and a unique function* $\hat{\theta}(x)$ *from* R_s *to* R_k *continuous in* N, *such that* $\hat{\theta}(P(\theta)) = \theta$ *for* $\theta \varepsilon \Theta$, *and* $F(x, \hat{\theta}(x)) = 0$ *for* $x \varepsilon N$. *Moreover,* $\mathfrak{L}\{\sqrt{n}(\hat{\theta}(Z_n) - \theta) | \theta\} \to \mathfrak{L}(Y)$ *where* Y *is a normal random vector with mean zero and variance-covariance matrix given by*

$$[b(\theta)\dot{P}(\theta)']^{-1}b(\theta)\Sigma(\theta)b(\theta)'[\dot{P}(\theta)b(\theta)']^{-1}$$

PROOF. $F(P(\theta), \theta) = 0$ and

$$F_\theta(x, \theta) = f_\theta(x, \theta)(x - P(\theta)) - \dot{P}(\theta)f(x, \theta)' \tag{3.5}$$

where $f_\theta(x, \theta)$ represents the $k \times k \times s$ cubic matrix of partial derivatives of the $k \times s$ matrix $f(x, \theta)$ with respect to θ. To avoid confusion we will write out the first term of this difference completely. Denote the function in the ith row, jth column of $f(x, \theta)$ by $f_{ij}(x, \theta)$, and let $P_j(\theta)$ and x_j represent the jth component of the vectors $P(\theta)$ and x. Then,

$$f_\theta(x, \theta)(x - P(\theta)) = \sum_{j=1}^{s} \begin{pmatrix} \frac{\partial}{\partial\theta_1} f_{1j} & \cdots & \frac{\partial}{\partial\theta_1} f_{sj} \\ \vdots & & \\ \frac{\partial}{\partial\theta_k} f_{1j} & \cdots & \frac{\partial}{\partial\theta_k} f_{sj} \end{pmatrix} (x_j - P_j(\theta)). \tag{3.6}$$

It is now easily checked that formula (3.5) holds. Hence,

$$F_\theta(P(\theta), \theta) = -\dot{P}(\theta)b(\theta)' \tag{3.7}$$

which, by assumption, is nonsingular for every $\theta \varepsilon \Theta$. Thus the hypotheses of the lemma of the previous section are satisfied and the first part of the theorem is proved.

To prove the second part, expand $F(x, \theta)$ about the point $\hat{\theta}(x)$ to one term using the formula

$$F(x, \theta) = F(x, \hat{\theta}(x)) + \left[\int_0^1 F_\theta\{x, \hat{\theta}(x) + \lambda(\theta - \hat{\theta}(x)\} \, d\lambda\right]' (\theta - \hat{\theta}(x)) \tag{3.8}$$

which may easily be checked. By the integral of a matrix we mean the matrix of the integrals of each term separately. For each $\theta \varepsilon \Theta$, formula (3.8) is valid whenever x is sufficiently close to $p(\theta)$, so that $(x, \hat{\theta}(x))$ is in a spherical neighborhood of $(p(\theta), \theta)$ contained entirely in N_0. We may replace x by Z_n in (3.8) and multiply both sides by $\sqrt{n}$.

$$\begin{aligned}\sqrt{n}\left[-\int_0^1 F_\theta\{Z_n, \hat{\theta}(Z_n) + \lambda(\theta - \hat{\theta}(Z_n))\}\, d\lambda\right]' (\hat{\theta}(Z_n) - \theta) \\ = f(Z_n, \theta)\sqrt{n}\,(Z_n - P(\theta)).\end{aligned} \tag{3.9}$$

We now invoke the theorems of Slutsky (see [10], section 2, theorem 2, or [4]). From assumption 1, $\mathfrak{L}(Z_n \mid \theta) \to \mathfrak{L}(P(\theta))$. Hence by Slutsky's theorem, since $f(x, \theta)$ is continuous in a neighborhood of $(p(\theta), \theta)$,

$$\mathfrak{L}(f(Z_n, \theta) \mid \theta) \to \mathfrak{L}(f(P(\theta), \theta)) = \mathfrak{L}(b(\theta)). \tag{3.10}$$

Slutsky's theorem also gives

$$\mathfrak{L}(f(Z_n, \theta)\sqrt{n}(Z_n - P(\theta)) \mid \theta) \to \mathfrak{L}(b(\theta)Z) \tag{3.11}$$

where Z is a normal vector with zero mean and variance-covariance matrix $\Sigma(\theta)$. Since $\mathfrak{L}(Z_n \mid \theta) \to \mathfrak{L}(P(\theta))$ and $\mathfrak{L}(\hat{\theta}(Z_n) \mid \theta) \to \mathfrak{L}(\hat{\theta}(P(\theta))) = \mathfrak{L}(\theta)$, we may apply the Lebesgue bounded convergence theorem to the integral in (3.9).

$$\begin{aligned}\mathfrak{L}\left\{\int_0^1 F_\theta[Z_n, \hat{\theta}(Z_n) + \lambda(\theta - \hat{\theta}(Z_n))]\, d\lambda \mid \theta\right\} \to \mathfrak{L}\left\{\int_0^1 F_\theta[P(\theta), \theta]\, d\lambda\right\} \\ = \mathfrak{L}\{F_\theta(P(\theta), \theta)\} = \mathfrak{L}\{-\dot{P}(\theta)b(\theta)'\}\end{aligned} \tag{3.12}$$

by equation (3.7). Another application of Slutsky's theorem allows us to conclude

$$\mathfrak{L}\{\sqrt{n}(\hat{\theta}(Z_n) - \theta) \mid \theta\} \to \mathfrak{L}\{[b(\theta)\dot{P}(\theta)']^{-1}b(\theta)Z\}. \tag{3.13}$$

Denoting $[b(\theta)\dot{P}(\theta)']^{-1}b(\theta)Z$ by Y, we see that Y is a normal random vector, with mean zero and covariance matrix

$$\begin{aligned}EYY' &= E[b(\theta)\dot{P}(\theta)']^{-1}b(\theta)ZZ'b(\theta)'[\dot{P}(\theta)b(\theta)']^{-1} \\ &= [b(\theta)\dot{P}(\theta)']^{-1}b(\theta)\Sigma(\theta)b(\theta)'[\dot{P}(\theta)b(\theta)']^{-1}.\end{aligned} \tag{3.14}$$

3.3. Applications. The theorem just proved allows some immediate inferences. The important point to notice in this theorem is that the asymptotic distribution of $\sqrt{n}(\hat{\theta}(Z_n) - \theta)$ depends on the function $f(x, \theta)$ only through its values along the curve $\{(P(\theta), \theta) : \theta \varepsilon \Theta\}$. Thus if the linear form

$$F(Z_n, \theta) = f(Z_n, \theta)(Z_n - P(\theta))$$

has a root which is already a B.A.N. estimate of θ, any linear form

$$g(Z_n, \theta)(Z_n - P(\theta)),$$

in which the function $f(x, \theta)$ is replaced by any function $g(x, \theta)$ satisfying assumption 4 and for which $g(P(\theta), \theta) = f(P(\theta), \theta)$, will have a root which is also a B.A.N. estimate of θ, since the asymptotic distribution of the two roots will be the same.

For example, equation (3.2) (neglecting the factor n which is immaterial as far as roots are concerned) is a linear form of the type $f(Z_n, \theta)(Z_n - P(\theta))$ for which

$$f(Z_n, \theta) = \dot{P}(\theta)M(Z_n). \tag{3.15}$$

Since $M(Z_n)$ converges in probability to $\Sigma(\theta)^{-1}$ when θ is the true value of the parameter, $M(P(\theta)) = \Sigma(\theta)^{-1}$ so that

$$b(\theta) = \dot{P}(\theta)\Sigma(\theta)^{-1} \tag{3.16}$$

Now consider functions

$$f_1(Z_n, \theta) = b(\theta) \quad \text{and} \quad f_2(Z_n, \theta) = L(Z_n)M(Z_n) \tag{3.17}$$

where L is a matrix continuous in a neighborhood of $\{P(\theta): \theta \varepsilon \Theta\}$, such that $L(P(\theta)) = \dot{P}(\theta)$. If $f_1(Z_n, \theta)$ is used, we must also assume that $b(\theta)$ has a continuous derivative. In these circumstances, whenever the root to equation (3.2) is a B.A.N. estimate, roots to the linear forms involving $f_1(Z_n, \theta)$ and $f_2(Z_n, \theta)$ will be B.A.N. also.

Now we will show directly the exact conditions under which there will be a root of a linear form which will be "best" out of the class of all roots of linear forms; that is, the exact conditions under which there is a value of $b(\theta)$ which minimizes the variance (3.14).

Of two n by n matrices, A and B, A will be said to be smaller than B, in symbols $A < B$, if and only if $B - A$ is positive semi-definite; that is, if

$$x'[B - A]x \geqq 0$$

for every n-dimensional vector x. Thus of two unbiased estimates of a vector parameter θ, T_1 and T_2, with covariance matrices respectively A and B, T_1 would be preferred to T_2 if $A < B$, since the unbiased estimate $x'T_1$ of the parameter $x'\theta$ will have a smaller variance than the unbiased estimate $x'T_2$ of the same parameter.

THEOREM 2. *If in addition to assumptions* 1 *through* 5 *there exists an* s *by* s *nonsingular matrix* $\Sigma_0(\theta)$ *such that*

$$\Sigma(\theta)\Sigma_0(\theta)\dot{P}(\theta)' = \dot{P}(\theta)' \tag{3.18}$$

then the asymptotic covariance matrix of $\hat{\theta}(Z_n)$ *taken on its minimum value when* $b(\theta) = \dot{P}(\theta)\Sigma_0(\theta)$. *The minimum value is then* $[\dot{P}(\theta)\Sigma_0(\theta)\dot{P}(\theta)']^{-1}$.

PROOF. For simplicity of notation the θ will be omitted. From assumption 5, $\dot{P}$ is of full rank so that $[\dot{P}\Sigma_0\dot{P}]$ is nonsingular. The inequality

$$(b'[\dot{P}b']^{-1} - \Sigma_0\dot{P}'[\dot{P}\Sigma_0\dot{P}']^{-1})\Sigma(b[\dot{P}b']^{-1} - \Sigma_0\dot{P}'[\dot{P}\Sigma_0\dot{P}]^{-1}) \geqq 0 \tag{3.19}$$

which holds since Σ is positive semi-definite, yields

$$[b\dot{P}']^{-1}b\Sigma b'[\dot{P}b']^{-1} - [\dot{P}\Sigma_0\dot{P}']^{-1} \geqq 0. \tag{3.20}$$

Yet it is easily checked that equality is attained if $b = \dot{P}\Sigma_0'$. qed.

The assumption of the existence of a matrix $\Sigma_0(\theta)$ satisfying (3.18) holds for example when $\Sigma(\theta)$ is nonsingular. Then $b(\theta) = \dot{P}(\theta)\Sigma(\theta)^{-1}$ as was found in equation (3.16). However, in other important cases, for example in the multinomial case with the χ^2 of equation (2.2), the matrix $\Sigma(\theta)$ is singular. The following lemma which may be proved without difficulty, will perhaps be of aid in checking whether a Σ_0 satisfying (3.18) exists at all.

LEMMA. *In order that there exist a nonsingular matrix $\Sigma_0(\theta)$ satisfying* (3.18), *it is necessary and sufficient that the range space of $\dot{P}(\theta)'$ be contained in the range space of $\Sigma(\theta)$: that is, for every vector x there exists a vector $y(\theta)$ such that*

$$\Sigma(\theta)y(\theta) = \dot{P}(\theta)'x.$$

In certain cases one can find the matrix Σ_0 which satisfies (3.18). We shall do it now for the multinomial case. In this case the vector $P(\theta)$ is simply the vector of cell probabilities, and is $s + 1$ dimensional. The matrix $\Sigma(\theta)$ is found to be

$$\Sigma(\theta) = \begin{pmatrix} p_1(\theta) - p_1^2(\theta) & -p_1(\theta)p_2(\theta) & \cdots & -p_1(\theta)p_{s+1}(\theta) \\ -p_1(\theta)p_2(\theta) & p_2(\theta) - p_2^2(\theta) & & \\ \vdots & & & \\ -p(\theta)p_{s+1}(\theta) & & \cdots & p_{s+1}(\theta) - p_{s+1}^2(\theta) \end{pmatrix} \tag{3.21}$$

which may be expressed simply as

$$\Sigma(\theta) = B(\theta) - P(\theta)P(\theta)' \tag{3.22}$$

where $B(\theta)$ is the diagonal matrix

$$B(\theta) = \begin{pmatrix} p_1(\theta) & 0 & \cdots & 0 \\ 0 & p_2(\theta) & & \\ \cdot & & \cdot & \\ \cdot & & \cdot & \\ \cdot & & \cdot & \\ 0 & & & p_{s+1}(\theta) \end{pmatrix} \tag{3.23}$$

Then, as suggested by the χ^2 of (2.2), put $\Sigma_0(\theta) = B(\theta)^{-1}$.

$$\Sigma(\theta)\Sigma_0(\theta)\dot{P}(\theta)' = B(\theta)B(\theta)^{-1}\dot{P}(\theta)' - P(\theta)P(\theta)'B(\theta)^{-1}\dot{P}(\theta)'. \tag{3.24}$$

It is easily seen that

$$P(\theta)'B(\theta)^{-1}\dot{P}(\theta)' = \left(\sum_{i=1}^{s+1} \frac{\partial}{\partial\theta_1} p_i(\theta), \sum_{i=1}^{s+1} \frac{\partial}{\partial\theta_2} p_i(\theta), \cdots, \sum_{i=1}^{s+1} \frac{\partial}{\partial\theta_k} p_i(\theta)\right). \tag{3.25}$$

This vector must be zero since $\sum_{i=1}^{s+1} p_i(\theta) = 1$. Hence, the equality (3.18) is satisfied. Thus applying Theorem 2, roots of the linear form

$$\sum_{i=1}^{s+1} (z_i - p_i(\theta))f_{ij}(s_1, \cdots, s_{s+1}, \theta) = 0 \qquad j = 1, 2, \cdots k, \tag{3.26}$$

will be "best" when $f_{ij}(p_1(\theta), \cdots, p_{s+1}(\theta), \theta) = \partial/\partial\theta_j \log p_i(\theta)$.

It may further be shown in the multinomial case, that if the $f_{ij}(z, \theta)$ are chosen to be independent of z, and equal to $\partial/\partial\theta_j \log p_i(\theta)$, equation (3.26) will b the derivative of the log of the likelihood function set equal to zero, so that one has immediately that the maximum likelihood estimate, in addition to the minimum modified χ^2 estimate, is an estimate which is given by the root of a certain linear form. One would expect that the linear form (3.26) in which the functions f_{ij} do not depend on θ at all would be somewhat easier to solve for θ. It is this type of linear form which is suggested in section 4 as a method for estimating the bacterial density in a liquid.

We will now apply the preceding theorem to the various minimum χ^2 methods discussed previously.

Application to the transformed χ^2. The method of generating B.A.N. estimates described in Theorems 1 and 2 also applies easily to the transformed χ^2 of equations (2.8) and (2.10). For example, the derivative of the χ^2 of equation (2.8) with $T(Z_n)$ depending on Z_n only, and not on θ, is found to be

$$\frac{\partial}{\partial\theta}\chi_T^2 = n\dot{P}(\theta)\dot{g}(P(\theta))T(Z_n)(g(Z_n) - g(P(\theta))). \tag{3.27}$$

Assumption 1 of Theorem 1 becomes in this case

$$\mathcal{L}\{\sqrt{n}[g(Z_n) - g(P(\theta))] \mid \theta\} \to \mathcal{L}(Z) \tag{3.28}$$

where Z is a normal random vector with zero mean and variance-covariance matrix $[\dot{g}(P(\theta))\Sigma(\theta)\dot{g}(P(\theta))']$. This may easily be checked by expanding $g(Z_n)$ in a Taylor series about the point $P(\theta)$, and invoking asymptotic normality of $\sqrt{n}(Z_n - P(\theta))$. The only requirement on the function $g(x)$ is that it have a continuous derivative in a neighborhood of the curve $\{P(\theta):\theta \in \Theta\}$. If in addition $g(P(\theta))$ is nonsingular for each $\theta \in \Theta$, $[\dot{g}(P(\theta))\Sigma(\theta)\dot{g}(P(\theta))]^{-1}$ will exist and $b(\theta)$ is found to be

$$b(\theta) = \dot{P}(\theta)\dot{g}(P(\theta))[\dot{g}(P(\theta))\Sigma(\theta)\dot{g}(P(\theta))']^{-1}. \tag{3.29}$$

Thus, if the root to equation (3.27) is a B.A.N. estimate, the root to the linear form

$$f(Z_n, \theta)(g(Z_n) - g(P(\theta)) = 0 \tag{3.30}$$

will also be a B.A.N. estimate, provided that f satisfies Assumption 4, and that $f(P(\theta), \theta) = b(\theta)$.

The linear form corresponding to the transformed multinomial χ^2 of (2.10) may be computed as before. It becomes

$$\sum_{i=1}^{s+1} [g_i(z_i) - g_i(p_i(\theta))]f_{ij}(z_1, \cdots, z_{s+1}, \theta) = 0 \qquad j = 1, 2, \cdots, k \tag{3.31}$$

where

$$f_{ij}(p_1(\theta), \cdots, p_{s+1}(\theta), \theta) = \left[\frac{\partial}{\partial\theta_j} p_i(\theta)\right] \frac{1}{p_i(\theta)g_i'(p_i(\theta))}. \tag{3.32}$$

Under assumptions 1 through 5, and the assumptions that each $g_i(x)$ is continuous in a neighborhood of the curve $\{x:x = p_i(\theta), \theta \varepsilon \Theta\}$ and that

$$g_i'(p_i(\theta)) \neq 0,$$

the roots to equation (3.31) will be B.A.N. estimates of the parameters.

Application to the expansion of χ^2 *in a Taylor series.* Let θ_n^* be a $O(\sqrt{n})$-consistent estimate of the parameter θ. To find the minimum value of the right hand side of equation (2.12) without the remainder term, we take a derivative and solve for the root $\hat{\theta}$.

$$\hat{\theta}_n = \theta_n^* - \ddot{\chi}^2(\theta_n)^{-1}\dot{\chi}^2(\theta_n) \tag{3.33}$$

If we use the modified χ^2 of equation (2.3) for this procedure with M a function of Z_n only, for example $M(Z_n) = \Sigma(\theta_n^*)^{-1}$, the first two derivatives are

$$\begin{aligned} \dot{\chi}^2(\theta) &= 2n\dot{P}(\theta)\Sigma(\theta_n^*)^{-1}(Z_n - P(\theta)) \\ \ddot{\chi}^2(\theta) &= 2n\dot{P}(\theta)\Sigma(\theta_n^*)^{-1}\dot{P}(\theta)' - 2n\ddot{P}(\theta)\Sigma(\theta_n^*)^{-1}(Z_n - P(\theta)). \end{aligned} \tag{3.34}$$

where $\ddot{P}(\theta)$ is the $k \times k \times s$ cubic matrix of second partial derivatives of the vector $P(\theta)$.

If, on the other hand, we take the linear form with the function $f(Z_n, \theta)$ not depending on θ, say to be $\dot{P}(\theta_n^*)\Sigma(\theta_n^*)^{-1}$, and expand it about θ_n^* to the first power and solve for θ, we have

$$\hat{\theta}_n = \theta_n^* + [\dot{P}(\theta_n^*)\Sigma(\theta_n^*)^{-1}\dot{P}(\theta_n^*)]^{-1}\dot{P}(\theta_n^*)\Sigma(\theta_n^*)^{-1}(Z_n - P(\theta_n^*)). \tag{3.35}$$

If one compares the estimates (3.35) with the estimates (3.33) with equations (3.34) substituted, one sees that the former require less computation, and that by the amount in the second term of the expression for $\ddot{\chi}^2(\theta)$, involving all the second partial derivatives of the vector $P(\theta)$. Furthermore, computation of $[\dot{P}(\theta_n^*)\Sigma(\theta_n^*)^{-1}\dot{P}(\theta_n^*)]^{-1}$ would give an estimate of the limiting variance-covariance matrix of the B.A.N. estimate $\hat{\theta}_n$.

This method would be good for example in estimating the parameters of a Neyman type *A* distribution, where the vector $P(\theta)$ is a rather complicated function of the parameters, and other methods of getting B.A.N. estimates are rather difficult. This method has been applied by Robert Read of the Statistical Laboratory of the University of California, to estimating the parameters in a probabilistic model describing ionization in a cloud chamber, using as the preliminary estimates, those given by the method of moments. It has also been applied by Dr. Irene Rosenthal of the Psychology Department at the University of California, to estimate the parameters of a latent structure, using as first estimates those of Lazarsfeld [11].

4. Application to the problem of estimating bacterial density by the dilution method. The method of estimating the bacterial density of a liquid by taking samples in fermentation tubes at several levels of dilution of the liquid is well known. As far back as 1915 [12] the maximum likelihood estimate, called the

most probable number (M.P.N.) by Biometricians, was suggested for the problem, and is still being used today in Public Health for water, milk, and sewage analysis. This and other estimates have been studied by Fisher [13], Halvorson and Ziegler [14], and Matuszewski, Neyman, and Supinska [15].

The situation is the following. We are given a large volume V of a liquid containing a large number N of bacteria, and we are interested in estimating the bacterial density $\lambda = N/V$, the number of bacteria per unit volume. A sample of size α unit volume is withdrawn and tested by some device such as placing the sample in a fermentation tube to see if any bacteria are present. It is assumed that each bacterium acts independently and that each has the same probability α/V of being in the sample. Thus the number of bacteria in the sample will be binomially distributed with probability α/V and size N; however, if α/V is small and N is large the distribution may conveniently be replaced by a Poisson with parameter $N\alpha/V = \alpha\lambda$. The probability that no bacteria appear in the sample is then $p = e^{-\alpha\lambda}$. If n independent samples of size α are withdrawn and tested, the number K of sterile samples will be binomially distributed with probability p and size n, and may be used to estimate the parameter λ. However, the value of the experiment depends to a great extent on choosing α so that $p = e^{-\alpha\lambda}$ will be in a good estimating range, for if p is too small or too close to one, one will obtain too many fertile or too many sterile samples to be able to estimate λ with much accuracy. And since λ is unknown it will usually be impossible to choose α so that $e^{-\alpha\lambda}$ will be moderately between zero and one. So one usually takes several sizes of sample volumes α_1, $\alpha_2, \cdots, \alpha_s$, called dilution levels, and numbers of samples $n_1, n_2, \cdots, n_s$ at each of the levels, with the hope that at least one of the $e^{-\alpha_i\lambda}$ will be in a good estimating range. Then the numbers $k_1, k_2, \cdots, k_s$, of sterile samples at each of the levels will be used to estimate λ.

The most frequently used B.A.N. estimate of the bacterial density seems to be the maximum likelihood estimate, since the minimum χ^2 estimates appear to be much more difficult to compute. The maximum likelihood estimate of λ is that value of λ which is a root of the equation

$$\sum_{i=1}^{s} \frac{(n_i - k_i)\alpha_i}{(1 - e^{-\alpha_i\lambda})} = \sum_{i=1}^{s} n_i\alpha_i. \tag{4.1}$$

Methods of solving this equation have been discussed by Halvorson and Zeigler [14], Barkworth and Irwin [16], and Finney [17]. Tables of the estimate for certain situations may be found in Halvorson and Zeigler and in Hoskins [18].

An application of the methods of the previous section will yield a B.A.N. estimate which is slightly easier to compute. Linear forms which lead to B.A.N. estimates are of the type

$$\sum_{i=1}^{s} n_i f_i(z, \lambda)(z_i - e^{-\alpha_i\lambda}) \tag{4.2}$$

where z_i represents the frequency of sterile tubes at the ith level of dilution, $z_i = k_i/n_i$, and $f_i(z, \lambda)$ converges in probability to $\alpha_i(1 - e^{-\alpha_i\lambda})^{-1}$, z representing the vector $(z_1, \cdots, z_s)$. Equation (4.2) with $f_i(z, \lambda)$ always equal to

$$\alpha_i(1 - e^{-\alpha_i\lambda})^{-1}$$

is equivalent to the maximum likelihood equation (4.1).

We would like to replace $f_i(z, \lambda)$ in equation (4.2) completely by an estimate, that is, $f_i(z, \lambda) = \alpha_i/(1 - z_i)$, but we must take care of the cases in which z_i is equal to one. So we may choose $f_i(z, \lambda) = \alpha_i/(1 - z_i)$ if $z_i \neq 1$ and

$$f_i(z, \lambda) = \alpha_i(1 - e^{-\alpha_i\lambda})^{-1}$$

if $z_i = 1$. This will lead to a B.A.N. estimate since eventually as the n_i get large without bound, all the z_i will be different from one. We have the equation

$$\sum_{z_i \neq 1} n_i \frac{\alpha_i}{1 - z_i}(z_i - e^{-\alpha_i\lambda}) + \sum_{z_i = 1} n_i\alpha_i = 0. \tag{4.3}$$

Written in simpler form, this equation becomes

$$\sum_{z_i \neq 1} n_i \frac{\alpha_i}{1 - z_i} e^{-\alpha_i\lambda} = \sum_{z_i \neq 1} n_i \frac{\alpha_i}{1 - z_i} z_i + \sum_{z_i = 1} n_i \alpha_i . \tag{4.4}$$

This equation is simpler to solve than equation (4.1) in that it only requires tables of e^{-x} which are readily available, while equation (4.1) requires for its solution the computation of $(1 - e^{-\alpha_i\lambda})^{-1}$ separately for each i or tables of $(e^x - 1)^{-1}$ or $(1 - e^{-x})^{-1}$. The method by which it is suggested that (4.4) be solved is the same as that suggested by other authors in connection with the solution of (4.1), and that is Newton's method. For a function $f(x)$ with a continuous first derivative, if x_0 is taken to be the initial guess at the solution of $f(x) = 0$, x_n is defined inductively by

$$x_n = x_{n-1} - \frac{f(x_{n-1})}{f'(x_{n-1})}. \tag{4.5}$$

Applying this procedure to equation (4.4), we obtain the inductive formula

$$\lambda_n = \lambda_{n-1} + \frac{\displaystyle\sum_{z_i \neq 1} \frac{n_i\,\alpha_i}{1 - z_i} e^{-\alpha_i\lambda_{n-1}} - \sum_{z_i \neq 1} \frac{n_i\,\alpha_i}{1 - z_i} z_i - \sum_{z_i = 1} n_i\,\alpha_i}{\displaystyle\sum_{z_i \neq 1} \frac{n_i\,\alpha_i^2}{1 - z_i} e^{-\alpha_i\lambda_{n-1}}}. \tag{4.6}$$

The author has made a numerical study of the small sample properties of this estimate, the minimum χ^2 estimate and the maximum likelihood estimate, which he hopes to publish at a later date. An indication is given in this study that in general the estimate given by equation (4.4) has slightly better small sample properties in the sense of bias and root mean square error, than either the maximum likelihood or the minimum χ^2 estimate.

In conclusion, I would like to express my thanks to Professor L. Le Cam for his generous advice and helpful discussions concerning this paper.

REFERENCES

[1] J. NEYMAN, "Contribution to the theory of the χ^2 test," *Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press (1949).

[2] E. BARANKIN AND J. GURLAND, "On asymptotically normal efficient estimators: I," *University of California Publications in Statistics*, Vol. 1, No. 6 (1951), pp. 86–130.

[3] C. L. CHIANG, "On regular best asymptotically normal estimates with an application to a stochastic process," Doctoral Dissertation, University of California (1952).

[4] L. LE CAM, Lecture notes on large sample theory (mimeographed), University of California, 1955.

[5] W. TAYLOR, "Distance functions and regular best asymptotically normal estimates," *Ann. Math. Stat.*, Vol. 24, No. 1 (1953), pp. 85–92.

[6] J. BERKSON, "Application of the logistic function to bio-assay," *J. Amer. Stat. Assn.*, Vol. 39 (1944), pp. 357–365.

[7] H. CRAMÉR, *Mathematical Methods of Statistics*, Princeton University Press (1946).

[8] J. PIERPONT, *The Theory of Functions of Real Variables*, Vol. I, Ginn and Co., (1905).

[9] L. M. GRAVES, *The Theory of Functions of Real Variables*, McGraw-Hill Book Co., Inc. (1946).

[10] H. CHERNOFF, "Large sample theory: parametric case," *Ann. Math. Stat.*, Vol. 27, No. 1 (1956), pp. 1–22.

[11] S. A. STOUFFER, *Measurement and Prediction*, Princeton University Press (1950).

[12] M. H. McCRADY, "The numerical interpolation of fermentation-tube results," *Jour. Infec. Dis.*, Vol. 17 (1915), pp. 183–212.

[13] R. A. FISHER, "On the mathematical foundations of theoretical statistics," *Phil. Trans. Roy. Soc., London, Ser. A.*, Vol. 222 (1922), pp. 309–368.

[14] H. O. HALVORSEN AND N. R. ZIEGLER, "Application of statistics to problems in bacteriology," *Jour. Bact.*, Vol. 25 (1933), pp. 101–121.

[15] T. MATUSZEWSKI, J. NEYMAN, AND J. SUPINSKA, "Statistical studies in questions of bacteriology," *J. Roy. Stat. Soc., Ser. B.*, Vol. 2 (1932), pp. 63–82.

[16] H. BARKSWORTH AND J. O. IRWIN, "Distribution of coliform organisms in milk and the accuracy of the presumptive coliform test," *Jour. Hyg.*, Cambridge, Vol. 38 (1938), pp. 446–457.

[17] D. J. FINNEY, "The principle of biological assay," *J. Roy. Stat. Soc., Ser. B.*, Vol. 9 (1947), pp. 46–91.

[18] J. K. HOSKINS, "Most probable numbers for evaluation of coli aerogenes tests by fermentation tube method," Public Health Reports, Vol. 49 (1934), p. 393, Reprint No. 1621 (revised 1940).

FAMILIES OF DESIGNS FOR TWO SUCCESSIVE EXPERIMENTS

By G. H. Freeman

East Malling Research Station, Maidstone, England

It is sometimes desirable, particularly in experimentation with perennial plants, to conduct an experiment on plots already used for a previous trial. Various designs are known that facilitate this process (Hoblyn *et al.*, 1954), the following notation being used to describe types of design. The letters O, T and P refer respectively to designs that are orthogonal, totally balanced—i.e., balanced incomplete blocks—and partially balanced; then a design of type **X:YZ** where X, Y or Z may be any of O, T or P, is one in which the arrangement of the first set of treatments with respect to blocks is of type X, that of the second set of treatments to blocks is of type Y and that of the second set of treatments to the first is of type Z. It is assumed that the two sets of treatments are non-interacting and that designs of type T or P may be extended, i.e., have complete replicates in each block. Then, if the first trial is in randomised blocks, type O, the only design that has not previously been very fully discussed is type O:PP for which, however, general methods of analysis have been given (Freeman, 1957b). The purpose of this paper is to describe all known families of O:PP designs with two associate-classes. These, being designs with two orthogonal constraints, can also be regarded as row and column designs, and will henceforth be considered as such here.

The families of O:PP designs described here include all those with any members likely to be of much practical use, i.e., having more than two replicates or treatments, not more than 30 replicates, treatments, rows or columns and not more than 150 plots in all. The possibilities of existence of all O:PP designs within these limits have been investigated by enumeration and all the tabulated designs have been found to exist. Where larger designs are required their existence can usually be readily determined and, particularly where the number of replicates greatly exceeds the number of treatments, there may be many possible designs. A catalogue of the designs in Tables II–V and VII has been prepared, and is available at East Malling Research Station; the construction of an individual design gives rise to no practical difficulty by trial and error, but no attempt has been made to find transformation sets for each design.

Since, for practical purposes, O:PP designs are constrained to have the same associate-classes with respect to rows and columns their classification depends on that of designs of type P. These, partially balanced, designs have been described in great detail by Bose and his co-workers, who have provided an extensive catalogue of such designs with two associate-classes, (Bose *et al.*, 1954). Although this catalogue is now known not to be exhaustive (see, for example, Archbold and Johnson, 1956 and Freeman, 1957a) it does provide a basis for the

Received June 7, 1957.

classification of designs of type P and is thus adopted here for O:PP designs as well.

Group Divisible Designs

The simplest O:PP designs are those that are group divisible, the property of group divisibility being such that designs group divisible one way are also group divisible the other. Group divisible partially balanced designs can be divided into the three types, singular, semi-regular and regular, of which the first and last will be denoted as S[1] and R respectively, as usual, while semi-regular designs will be called H, so as to keep a one-letter code. As each type of group divisible design can be combined with itself or each other to make an O:PP design there are thus six types of group divisible O:PP design, these being described, in an obvious notation, as SS, HH, HS, RR, RH, RS, and considered in this order.

To classify group divisible O:PP designs into families it is first necessary to consider the various types of singular, semi-regular and regular designs.

Singular designs. Families of singular designs are uniquely determined by balanced incomplete block designs (Bose and Connor, 1952), and thus a complete classification of the former is afforded by a corresponding one of the latter. The following types of balanced incomplete block design are considered, these not including every possible design but containing all that give rise to singular designs from which can be constructed O:PP designs of practicable size. We shall use the notation $C[n; k]$ for the binomial coefficient $\binom{n}{k}$

(i) $C[v - 1; k - 1]$ replicates of v treatments on $C[v; k]$ blocks of k plots each (unreduced)
(ii) $(s + 1)$ replicates of s^2 treatments on $s(s + 1)$ blocks of s plots each (orthogonal series 1 or OS 1)
(iii) $(s^2 - 1)$ replicates of s^2 treatments on $s(s + 1)$ blocks of $s(s - 1)$ plots each (complement of OS 1)
(iv) $(2t + 1)$ replicates of $(2t + 2)$ treatments on $(4t + 2)$ blocks of $(t + 1)$ plots each
(v) $(2t + 1)$ replicates of $(4t + 3)$ treatments on $(4t + 3)$ blocks of $(2t + 1)$ plots each
(vi) $(2t + 2)$ replicates of $(4t + 3)$ treatments on $(4t + 3)$ blocks of $(2t + 2)$ plots each.

The first of these types exists for all values of v and all $k < v$, types (ii) and (iii) for all values of s for which complete sets of orthogonal Latin squares exist, and types (iv)–(vi) when $(4t + 3)$ is a prime-power (Bose, 1939); Bose also shows geometrically that, though $(4t + 3)$ is not a prime-power for $t = 3$, designs of

[1] The letters S here and T for triangular designs used later in the paper are also used respectively for designs with supplemented balance and total balance by Hoblyn *et al.* (1954), but there should be no confusion between the uses for *types* of balanced designs and *families* of partially balanced designs.

types (iv)–(vi) are possible for $t = 3$. Types (v) and (vi) include the second orthogonal series (OS 2) and its complement, respectively, when $t = 1$, the corresponding value of s in the orthogonal series designs being $s = 2$.

The classification of balanced incomplete block designs into these six types is not mutually exclusive; for example, the design with 3 replicates of 4 treatments on 6 blocks of 2 plots each can be considered as belonging to each of types (i), (ii), (iii) and (iv). It follows that particular O:PP designs singular one way may belong to more than one family. For the sake of uniqueness, any one design of type P or O:PP will be regarded as a member of only one family, since the overlapping of families is a direct consequence of the overlap of types of balanced incomplete block design and thus irrelevant to the consideration of O:PP designs.

A balanced incomplete block design with r^* replicates of v^* treatments on b^* blocks of k^* plots each gives an unextended singular P design with r^* replicates of nv^* treatments on b^* blocks with nk^* plots each, there being v^* groups of n treatments each, $n > 1$. Thus, on allowing extended designs with p complete replicates of treatments in each block and, further, the whole design repeated q times, the most general singular P has the following numbers of replicates, treatments, blocks and plots per block: $q(r^* + pb^*)$, nv^*, qb^*, $n(k^* + pv^*)$. The parameters of all designs of type P will be written in this order henceforth; m will be used instead of v^* so as to consider m groups of treatments in conformity with the usual notation.

The types of balanced incomplete block design enumerated above give rise to the following families of singular designs, where the inequalities are inserted, in S(ii)–(vi), to ensure the uniqueness of the families:

$$\text{S(i)} \qquad \frac{q(pm + k)}{m}\, C[m; k],\ mn,\ qC[m;k],\ n(pm + k) \qquad (k < m)$$

$$\text{S(ii)} \qquad q(s + 1)(ps + 1),\ ns^2,\ qs(s + 1),\ ns(ps + 1) \qquad (s > 2)$$

$$\text{S(iii)} \qquad q(s + 1)\,[s(p + 1) - 1],\ ns^2,\ qs(s + 1),\ ns[s(p + 1) - 1] \qquad (s > 2)$$

$$\text{S(iv)} \qquad q(2t + 1)(2p + 1),\ 2n(t + 1),\ 2q(2t + 1),\ n(t + 1)(2p + 1) \qquad (t > 1)$$

$$\text{S(v)} \qquad q[2t(2p + 1) + 3p + 1],\ n(4t + 3),\ q(4t + 3),\ n[2t(2p + 1) + 3p + 1] \qquad (t > 0)$$

$$\text{S(vi)} \qquad q[2t(2p + 1) + 3p + 2],\ n(4t + 3),\ q(4t + 3),\ n[2t(2p + 1) + 3p + 2] \qquad (t > 0)$$

Semi-regular designs. Bose *et al.* (1953) classify semi-regular designs according as λ_1 does or does not equal zero, but this classification seems unnecessary for the present purpose, $\lambda_1 = 0$ being merely a special case. Thus, from Bose *et al.*, the design has $n\lambda_2/c$ replicates of mn treatments in m groups of n on $n^2\lambda_2/c^2$ blocks of mc plots, where $\lambda_1 = n\lambda_2(c - 1)/c(n - 1)$ is integral and $m \leqq (n^2\lambda_2 - c^2)/c^2(n - 1)$. The extension of this design to allow for p complete replicates of the

treatments in each block and the whole design repeated q times gives the following family of semi-regular designs:

$$\text{H} \qquad \frac{qn\lambda_2(pn + c)}{c^2}, mn, \frac{qn^2\lambda_2}{c^2}, m(pn + c), \text{ where } n\, \frac{\lambda_2(c - 1)}{c(n - 1)}$$

is integral and

$$m \leqq \frac{n^2\lambda_2 - c^2}{c^2(n - 1)}$$

Regular designs. These are more difficult to categorise than either of the other types of group divisible design. Further, practicable O:PP designs cannot be derived from all types of regular design; two such types are those generated by the methods of differences and of omitting varieties (Bose *et al.*, 1953), and thus these two types are not considered further.

The types of regular design that are considered are as follows:

(i) designs derivable by addition,

(ii) designs with complete and incomplete groups,

(iii) designs with groups arranged in sets,

(iv) disconnected designs. The first of these consists of all designs derivable by addition of group divisible designs to other group divisible designs or to balanced incomplete block designs, while the next two types are described elsewhere (Freeman, 1957a). The fourth type is not considered in an unextended form, as it has been shown (Freeman, 1957c) that this type of design cannot give rise to an O:PP design; however, extended disconnected designs are of use for the construction of O:PP designs.

In order to consider only those designs of type P that give rise to O:PP designs when the plots within each block are rearranged in accordance with a second classification further restrictions on the parameters are necessary in types R(ii) and R(iv). With these further restrictions, the four families of regular designs are as follows, there being p complete replicates of the treatments in each block and the whole design being repeated q times:

$$\text{R(i)} \qquad \frac{qR(k + pmn)}{k}, mn, \frac{qRmn}{k}, k + pmn,$$

where

$R = ar + a'r', a, r, a', r' > 0, k > 1,$

$$\lambda_1(n - 1) + \lambda_2 n(m - 1) = r(k - 1), \lambda_1'(n - 1) + \lambda_2' n(m - 1) = r'(k - 1),$$

$$a\lambda_1 + a'\lambda_1' = \Lambda_1 \neq a\lambda_2 + a'\lambda_2' = \Lambda_2 .$$

$$\text{R(ii)} \qquad \frac{qC[m - 1; u]C[n; h](nu + h + pmn)}{n}, mn,$$

$$qmC[m - 1; u]C[n; h], nu + h + pmn,$$

where $0 < u < m$, $1 < h < n - 1$.

R(iii) $\quad q(3n - 1)(pn + 1), 2n^2, qn(3n - 1), 2n(pn + 1).$

R(iv) $\quad \dfrac{qr(k + pmn)}{k}, mn, \dfrac{qrmn}{k}, k + pmn,$

where $r(k - 1)/(n - 1)$ is integral, $p > 0$, $1 < k < n - 1$.

There are m groups of n treatments in each design, where, for family R(iii), $m = 2n$.

Construction of O:PP designs. Not every pair of designs of type P can give rise to an O:PP design. Thus, families S(ii) and S(iii) are incompatible with S(v) and S(vi), as s^2 is never of the form $(4t + 3)$. Also R(iii) is incompatible with R(ii) and R(iv) as, in order to satisfy the relations $qn(3n - 1) = nu + h + pmn$ or $qn(3n - 1) = k + pmn$, h or k must be a multiple of n, an impossibility since each lies between 1 and $(n - 1)$. Further, S(iv) and R(iii) require an even number of groups of treatments and so cannot be associated with S(v) and S(vi), which require an odd number.

All the families of group divisible O:PP designs not excluded by the above argument are given in Table I together with their derivation from the corresponding families of P designs. As an example, consider family SS II, derived from families S(ii) and S(i). For the numbers of groups of treatments to be the same in the two designs the relation $m = s^2$ must be satisfied, while for the blocks and plots per block of the two families to be interchangeable two further relations must hold. Throughout Table I, to distinguish between the parameters in the two families of type P, that written second has dashes, and so here p' and q' refer to S(i) while p and q refer to S(ii). The relations between blocks and plots per block then are: $qs(s + 1) = n(p's^2 + k)$, $q'C[s^2; k] = ns(ps + 1)$. Thus

$$q = \frac{n(p's^2 + k)}{s(s + 1)} \text{ and } q' = \frac{ns(ps + 1)}{C[s^2; k]}$$

as shown in Table I. The number of treatments in the design is ns^2, while the last two columns of Table I show the numbers of blocks in the two designs of type P, i.e., $qs(s + 1)$ for S(ii) and $q'C[s^2; k]$ for S(i). The number of replicates is shown in Table I as $q(s + 1)(ps + 1)$, that corresponding to S(ii), although it could be given in several forms; by convention, the number is given throughout Table I in terms of the replicates of the family of type P given first.

In Table I, all the numbers shown are non-negative integers and are subject to the restrictions described in the classification above of designs of type P. In certain families further restrictions are necessary on the values of the parameters by virtue of the first of the non-existence theorems (Freeman, 1957c); thus in any family constructed from S(i) the number of treatments in the O:PP design must be greater than or equal to the number of rows plus the number of columns, while in family SS III p and p' cannot both be zero. In families SS VIII and SS IX, in order to satisfy the relation $s^2 = 2(t + 1)$, the parameter w is intro-

TABLE I

Families of Group Divisible O:PP Designs

Family	Derivation (a) x (b)	q	q'	Replicates	Treatments	Blocks (a)	Blocks (b)
SS I	S(i) x S(i)	$\frac{n(p'm + k')}{C[m;k]}$	$\frac{n(pm + k)}{C[m;k']}$	$\frac{q(pm + k)}{m} C[m;k]$	mn	qC[m;k]	q'C[m;k']
SS II	S(ii) x S(i)	$\frac{n(p's^2 + k)}{s(s + 1)}$	$\frac{ns(ps + 1)}{C[s^2;k]}$	q(s + 1)(ps + 1)	ns^2	qs(s + 1)	$q'C[s^2;k]$
SS III	S(ii) x S(ii)	$\frac{n(p's + 1)}{s + 1}$	$\frac{n(ps + 1)}{s + 1}$	q(s + 1)(ps + 1)	ns^2	qs(s + 1)	q's(s + 1)
SS IV	S(iii) x S(i)	$\frac{n(p's^2 + k)}{s(s + 1)}$	$\frac{ns[s(p + 1) - 1]}{C[s^2;k]}$	q(s + 1)[s(p + 1) − 1]	ns^2	qs(s + 1)	$q'C[s^2;k]$
SS V	S(iii) x S(ii)	$\frac{n(p's + 1)}{s + 1}$	$\frac{n[s(p + 1) - 1]}{s + 1}$	q(s + 1)[s(p + 1) − 1]	ns^2	qs(s + 1)	q's(s + 1)
SS VI	S(iii) x S(iii)	$\frac{n[s(p' + 1) - 1]}{s + 1}$	$\frac{n[s(p + 1) - 1]}{s + 1}$	q(s + 1)[s(p + 1) − 1]	ns^2	qs(s + 1)	q's(s + 1)
SS VII	S(iv) x S(i)	$\frac{n[2p'(t + 1) + k]}{2(2t + 1)}$	$\frac{n(t + 1)(2p + 1)}{C[2(t + 1);k]}$	q(2t + 1)(2p + 1)	2n(t + 1)	2q(2t + 1)	q'C[2(t + 1);k]
SS VIII	S(iv) x S(ii)	$\frac{nw(2p'w + 1)}{4w^2 - 1}$	$\frac{nw(2p + 1)}{2w + 1}$	$q(4w^2 - 1)(2p + 1)$	$4nw^2$	$2q(4w^2 - 1)$	2q'w(2w + 1)
SS IX	S(iv) x S(iii)	$\frac{nw[2w(p' + 1) - 1]}{4w^2 - 1}$	$\frac{nw(2p + 1)}{2w + 1}$	$q(4w^2 - 1)(2p + 1)$	$4nw^2$	$2q(4w^2 - 1)$	2q'w(2w + 1)
SS X	S(iv) x S(iv)	$\frac{n(t + 1)(2p' + 1)}{2(2t + 1)}$	$\frac{n(t + 1)(2p + 1)}{2(2t + 1)}$	q(2t + 1)(2p + 1)	2n(t + 1)	2q(2t + 1)	2q'(2t + 1)
SS XI	S(v) x S(i)	$\frac{n[p'(4t + 3) + k]}{4t + 3}$	$\frac{n[2t(2p + 1) + 3p + 1]}{C[4t + 3;k]}$	q[2t(2p + 1) + 3p + 1]	n(4t + 3)	q(4t + 3)	q'C[4t + 3;k]
SS XII	S(v) x S(v)	$\frac{n[2t(2p' + 1) + 3p' + 1]}{4t + 3}$	$\frac{n[2t(2p + 1) + 3p + 1]}{4t + 3}$	q[2t(2p + 1) + 3p + 1]	n(4t + 3)	q(4t + 3)	q'(4t + 3)

TABLE I (Continued)

Family	Derivation (a) x (b)	q	q'	Replicates	Treatments	Blocks (a)	Blocks (b)
SS XIII	S(vi) x S(i)	$\frac{n[p'(4t+3)+k]}{4t+3}$	$\frac{n[2t(2p+1)+3p+2]}{C[4t+3;k]}$	$q[2t(2p+1)+3p+2]$	$n(4t+3)$	$q(4t+3)$	$q'C[4t+3;k]$
SS XIV	S(vi) x S(v)	$\frac{n[2t(2p'+1)+3p'+1]}{4t+3}$	$\frac{n[2t(2p+1)+3p+2]}{4t+3}$	$q[2t(2p+1)+3p+2]$	$n(4t+3)$	$q(4t+3)$	$q'(4t+3)$
SS XV	S(vi) x S(vi)	$\frac{n[2t(2p'+1)+3p'+2]}{4t+3}$	$\frac{n[2t(2p+1)+3p+2]}{4t+3}$	$q[2t(2p+1)+3p+2]$	$n(4t+3)$	$q(4t+3)$	$q'(4t+3)$
HH	H x H	$\frac{mc^2(p'n+c')}{n^2\lambda_2}$	$\frac{mc'^2(pn+c)}{n^2\lambda_2'}$	$\frac{qn\lambda_2(pn+c)}{c^2}$	mn	$\frac{qn^2\lambda_2}{c^2}$	$\frac{q'n^2\lambda_2'}{c'^2}$
HS I	H x S(i)	$\frac{c^2(p'm+k)}{n\lambda_2}$	$\frac{m(pn+c)}{C[m;k]}$	$\frac{qn\lambda_2(pn+c)}{c^2}$	mn	$\frac{qn^2\lambda_2}{c^2}$	$q'C[m;k]$
HS II	H x S(ii)	$\frac{c^2s(p's+1)}{n\lambda_2}$	$\frac{s(pn+c)}{s+1}$	$\frac{qn\lambda_2(pn+c)}{c^2}$	ns^2	$\frac{qn^2\lambda_2}{c^2}$	$q's(s+1)$
HS III	H x S(iii)	$\frac{c^2s[s(p'+1)-1]}{n\lambda_2}$	$\frac{s(pn+c)}{s+1}$	$\frac{qn\lambda_2(pn+c)}{c^2}$	ns^2	$\frac{qn^2\lambda_2}{c^2}$	$q's(s+1)$
HS IV	H x S(iv)	$\frac{c^2(t+1)(2p'+1)}{n\lambda_2}$	$\frac{(t+1)(pn+c)}{2t+1}$	$\frac{qn\lambda_2(pn+c)}{c^2}$	$2n(t+1)$	$\frac{qn^2\lambda_2}{c^2}$	$2q'(2t+1)$
HS V	H x S(v)	$\frac{c^2[2t(2p'+1)+3p'+1]}{n\lambda_2}$	$pn+c$	$\frac{qn\lambda_2(pn+c)}{c^2}$	$n(4t+3)$	$\frac{qn^2\lambda_2}{c^2}$	$q'(4t+3)$
HS VI	H x S(vi)	$\frac{c^2[2t(2p'+1)+3p'+2]}{n\lambda_2}$	$pn+c$	$\frac{qn\lambda_2(pn+c)}{c^2}$	$n(4t+3)$	$\frac{qn^2\lambda_2}{c^2}$	$q'(4t+3)$

TABLE I (Continued)

Family	Derivation (a) × (b)	q	q'	Replicates	Treatments	Blocks (a)	Blocks (b)
RR I	R(i) × R(i)	$\frac{k(k' + p'mn)}{Rmn}$	$\frac{k'(k + pmn)}{R'mn}$	$\frac{qR(k + pmn)}{k}$	mn	$\frac{qRmn}{k}$	$\frac{q'R'mn}{k'}$
RR II	R(ii) × R(i)	$\frac{k + p'mn}{mC[m-1;u]C[n;h]}$	$\frac{k(nu + h + pmn)}{Rmn}$	$\frac{qC[m-1;u]C[n;h](nu + h + pmn)}{n}$	mn	$qmC[m-1;u]C[n;h]$	$\frac{q'Rmn}{k}$
RR III	R(ii) × R(ii)	$\frac{nu' + h' + p'mn}{mC[m-1;u]C[n;h]}$	$\frac{nu + h + pmn}{mC[m-1;u']C[n;h']}$	$\frac{qC[m-1;u]C[n;h](nu + h + pmn)}{n}$	mn	$qmC[m-1;u]C[n;h]$	$q'mC[m-1;u']C[n;h']$
RR IV	R(iii) × R(i)	$\frac{k + 2p'n^2}{n(3n - 1)}$	$\frac{k(pn + 1)}{Rn}$	$q(3n - 1)(pn + 1)$	$2n^2$	$qn(3n - 1)$	$\frac{2q'Rn^2}{k}$
RR V	R(iii) × R(iii)	$\frac{2(p'n + 1)}{3n - 1}$	$\frac{2(pn + 1)}{3n - 1}$	$q(3n - 1)(pn + 1)$	$2n^2$	$qn(3n - 1)$	$q'n(3n - 1)$
RR VI	R(iv) × R(i)	$\frac{k(k' + p'mn)}{rmn}$	$\frac{k'(k + pmn)}{Rmn}$	$\frac{qr(k + pmn)}{k}$	mn	$\frac{qrmn}{k}$	$\frac{q'Rmn}{k'}$
RR VII	R(iv) × R(ii)	$\frac{k(nu + h + p'mn)}{rmn}$	$\frac{k + pmn}{mC[m-1;u]C[n;h]}$	$\frac{qr(k + pmn)}{k}$	mn	$\frac{qrmn}{k}$	$q'mC[m-1;u]C[n;h]$
RR VIII	R(iv) × R(iv)	$\frac{k(k' + p'mn)}{rmn}$	$\frac{k'(k + pmn)}{r'mn}$	$\frac{qr(k + pmn)}{k}$	mn	$\frac{qrmn}{k}$	$\frac{q'r'mn}{k'}$
RH I	R(i) × H	$\frac{k(p'n + c)}{Rn}$	$\frac{c^2(k + pmn)}{n^2\lambda_2}$	$\frac{qR(k + pmn)}{k}$	mn	$\frac{qRmn}{k}$	$\frac{q'n^2\lambda_2}{c^2}$
RH II	R(ii) × H	$\frac{p'n + c}{C[m-1;u]C[n;h]}$	$\frac{c^2(nu + h + pmn)}{n^2\lambda_2}$	$\frac{qC[m-1;u]C[n;h](nu + h + pmn)}{n}$	mn	$qmC[m-1;u]C[n;h]$	$\frac{q'n^2\lambda_2}{c^2}$
RH III	R(iii) × H	$\frac{2(p'n + c)}{3n - 1}$	$\frac{2c^2(pn + 1)}{n\lambda_2}$	$q(3n - 1)(pn + 1)$	$2n^2$	$qn(3n - 1)$	$\frac{q'n^2\lambda_2}{c^2}$
RH IV	R(iv) × H	$\frac{k(p'n + c)}{rn}$	$\frac{c^2(k + pmn)}{n^2\lambda_2}$	$\frac{qr(k + pmn)}{k}$	mn	$\frac{qrmn}{k}$	$\frac{q'n^2\lambda_2}{c^2}$

TABLE I (Continued)

Family	Derivation (a) × (b)	q	q'	Replicates	Treatments	Blocks (a)	Blocks (b)
RS I	R(i) × S(i)	$\frac{k(p'm + k')}{Rm}$	$\frac{k + pmn}{C[m;k']}$	$\frac{qR(k + pmn)}{k}$	mn	$\frac{qRmn}{k}$	$q'C[m;k']$
RS II	R(i) × S(ii)	$\frac{k(p's + 1)}{Rs}$	$\frac{k + pns^2}{s(s + 1)}$	$\frac{qR(k + pns^2)}{k}$	ns^2	$\frac{qRns^2}{k}$	$q's(s + 1)$
RS III	R(i) × S(iii)	$\frac{k[s(p' + 1) - 1]}{Rs}$	$\frac{k + pns^2}{s(s + 1)}$	$\frac{qR(k + pns^2)}{k}$	ns^2	$\frac{qRns^2}{k}$	$q's(s + 1)$
RS IV	R(i) × S(iv)	$\frac{k(2p' + 1)}{2R}$	$\frac{k + 2pn(t + 1)}{2(2t + 1)}$	$\frac{qR[k + 2pn(t + 1)]}{k}$	$2n(t + 1)$	$\frac{2qRn(t + 1)}{k}$	$2q'(2t + 1)$
RS V	R(i) × S(v)	$\frac{k[2t(2p' + 1) + 3p' + 1]}{R(4t + 3)}$	$\frac{k + pn(4t + 3)}{4t + 3}$	$\frac{qR[k + pn(4t + 3)]}{k}$	$n(4t + 3)$	$\frac{qRn(4t + 3)}{k}$	$q'(4t + 3)$
RS VI	R(i) × S(vi)	$\frac{k[2t(2p' + 1) + 3p' + 2]}{R(4t + 3)}$	$\frac{k + pn(4t + 3)}{4t + 3}$	$\frac{qR[k + pn(4t + 3)]}{k}$	$n(4t + 3)$	$\frac{qRn(4t + 3)}{k}$	$q'(4t + 3)$
RS VII	R(ii) × S(i)	$\frac{n(p'm + k)}{mC[m-1;u]C[n;h]}$	$\frac{nu + h + pmn}{C[m;k]}$	$\frac{qC[m-1;u]C[n;h](nu + h + pmn)}{n}$	mn	$qmC[m-1;u]C[n;h]$	$q'C[m;k]$
RS VIII	R(ii) × S(ii)	$\frac{n(p's + 1)}{sC[s^2-1;u]C[n;h]}$	$\frac{nu + h + pns^2}{s(s + 1)}$	$\frac{qC[s^2-1;u]C[n;h](nu + h + pns^2)}{n}$	ns^2	$qs^2C[s^2-1;u]C[n;h]$	$q's(s + 1)$
RS IX	R(ii) × S(iii)	$\frac{n[s(p' + 1) - 1]}{sC[s^2-1;u]C[n;h]}$	$\frac{nu + h + pns^2}{s(s + 1)}$	$\frac{qC[s^2-1;u]C[n;h](nu + h + pns^2)}{n}$	ns^2	$qs^2C[s^2-1;u]C[n;h]$	$q's(s + 1)$
RS X	R(ii) × S(iv)	$\frac{n(2p' + 1)}{2C[2t + 1;u]C[n;h]}$	$\frac{nu + h + 2pn(t + 1)}{2(2t + 1)}$	$\frac{qC[2t+1;u]C[n;h][nu+h+2pn(t+1)]}{n}$	$2n(t+1)$	$2q(t+1)C[2t+1;u]C[n;h]$	$2q'(2t + 1)$
RS XI	R(ii) × S(v)	$\frac{n[2t(2p' + 1) + 3p' + 1]}{(4t + 3)C[2(2t + 1);u]C[n;h]}$	$\frac{nu + h + pn(4t + 3)}{4t + 3}$	$\frac{qC[4t+2;u]C[n;h][nu+h+pn(4t+3)]}{n}$	$n(4t+3)$	$q(4t+3)C[4t+2;u]C[n;h]$	$q'(4t + 3)$

TABLE I (Continued)

Family	Derivation (a) x (b)	q	q'	Replicates	Treatments	Blocks (a)	Blocks (b)
RS XII	R(ii) x S(vi)	$\frac{n[2t(2p'+1)+3p'+2]}{(4t+3)C[2(2t+1);u]C[n;h]}$	$\frac{nu+h+pn(4t+3)}{4t+3}$	$\frac{qC[4t+2;u]C[n;h][nu+h+pn(4t+3)]}{n}$	$n(4t+3)$	$q(4t+3)C[4t+2;u]C[n;h]$	$q'(4t+3)$
RS XIII	R(iii) x S(i)	$\frac{2p'n+k}{3n-1}$	$\frac{2n(pn+1)}{C[2n;k]}$	$q(3n-1)(pn+1)$	$2n^2$	$qn(3n-1)$	$q'C[2n;k]$
RS XIV	R(iii) x S(ii)	$\frac{2w(2p'w+1)}{6w^2-1}$	$\frac{2w(2pw^2+1)}{2w+1}$	$q(6w^2-1)(2pw^2+1)$	$8w^4$	$2qw^2(6w^2-1)$	$2q'w(2w+1)$
RS XV	R(iii) x S(iii)	$\frac{2w[2w(p'+1)-1]}{6w^2-1}$	$\frac{2w(2pw^2+1)}{2w+1}$	$q(6w^2-1)(2pw^2+1)$	$8w^4$	$2qw^2(6w^2-1)$	$2q'w(2w+1)$
RS XVI	R(iii) x S(iv)	$\frac{(t+1)(2p'+1)}{3t+2}$	$\frac{(t+1)[p(t+1)+1]}{2t+1}$	$q(3t+2)[p(t+1)+1]$	$2(t+1)^2$	$q(t+1)(3t+2)$	$2q'(2t+1)$
RS XVII	R(iv) x S(i)	$\frac{k(p'm+k')}{rm}$	$\frac{k+pmn}{C[m;k']}$	$\frac{qr(k+pmn)}{k}$	mn	$\frac{qrmn}{k}$	$q'C[m;k']$
RS XVIII	R(iv) x S(ii)	$\frac{k(p's+1)}{rs}$	$\frac{k+pns^2}{s(s+1)}$	$\frac{qr(k+pns^2)}{k}$	ns^2	$\frac{qrns^2}{k}$	$q's(s+1)$
RS XIX	R(iv) x S(iii)	$\frac{k[s(p'+1)-1]}{rs}$	$\frac{k+pns^2}{s(s+1)}$	$\frac{qr(k+pns^2)}{k}$	ns^2	$\frac{qrns^2}{k}$	$q's(s+1)$
RS XX	R(iv) x S(iv)	$\frac{k(2p'+1)}{2r}$	$\frac{k+2pn(t+1)}{2(2t+1)}$	$\frac{qr[k+2pn(t+1)]}{k}$	$2n(t+1)$	$\frac{2qrn(t+1)}{k}$	$2q'(2t+1)$
RS XXI	R(iv) x S(v)	$\frac{k[2t(2p'+1)+3p'+1]}{r(4t+3)}$	$\frac{k+pn(4t+3)}{4t+3}$	$\frac{qr[k+pn(4t+3)]}{k}$	$n(4t+3)$	$\frac{qrn(4t+3)}{k}$	$q'(4t+3)$
RS XXII	R(iv) x S(vi)	$\frac{k[2t(2p'+1)+3p'+2]}{r(4t+3)}$	$\frac{k+pn(4t+3)}{4t+3}$	$\frac{qr[k+pn(4t+3)]}{k}$	$n(4t+3)$	$\frac{qrn(4t+3)}{k}$	$q'(4t+3)$

duced, where $s = 2w$ and so $(t + 1) = 2w^2$; similarly, in RS XIV and RS XV to satisfy $s^2 = 2n$, w is again introduced, with $s = 2w$ and $n = 2w^2$. These uses of w are the only occasions where auxiliary symbols are necessary in this Table.

Useful group divisible designs. Tables II–V give the numerical values of the parameters of all useful group divisible designs. Where there is a design with the same parameters but simpler than O:PP the O:PP design is not included in the appropriate Table. In these Tables, which are derived by putting numerical values in Table I, "rows" and "columns" are substituted for the two types of "blocks" of the basic designs of type P. Rows and columns are chosen not to correspond to either type of block but merely so that there are never more rows than columns, this convention being a help to the writing down of the designs themselves.

It will be seen that by far the greatest number of useful designs occurs in the three families SS I, HS I and RS I, the designs from which are enumerated in Tables II, III and IV respectively. As no other family has more than five useful designs, the useful designs in all other families are given in Table V, in which many of the parameters are inapplicable in each line.

Only ten families are represented in Table V, and so Table VI has been constructed from Table I showing the smallest designs that are theoretically possible in the other 43 families that have no useful designs. Since even the smallest of these, those for SS II and RR VI, have 216 plots, while the largest, that for RS XII, has 8744736 plots, no attempt has been made to discover whether these designs do in fact exist. None of them appears to be excluded by any of the non-existence theorems (Freeman, 1957c), and so they are all presumed to be possible; however none is likely to be practicable. Table VI, like Table V, has many parameters inapplicable in each line.

In Tables IV, V and VI, whenever designs derived from the P design R(i) are given, the auxiliary parameters Λ_1 and Λ_2 are shown. This is because two designs, like RS I 1 and RS I 2 in Table IV, may differ from each other only in respect of these parameters. No other auxiliary parameters are necessary, as these are relevant only to the construction of the design, not its final form.

Non-Group Divisible Designs

Like designs of type P, O:PP designs that are group divisible are much more numerous than those that are not. Of the other types of P design, only the triangular and Latin square appear to give rise to O:PP designs; while the other types may, none of the designs given by Bose *et al.* (1954) leads to an O:PP design and, at least for cyclic designs, it appears unlikely that any could. Since in an O:PP design the association scheme is the same both ways designs that are either triangular or Latin square one way are the same the other, the association schemes being unique for each type.

Triangular designs. The basic triangular design has r replicates of $n(n - 1)/2$ treatments on $rn(n - 1)/2k$ blocks of k plots each. If $n = 2$ the design has only

TABLE II

Useful Designs in Family SS I

Design	m	n	k	k′	p	p′	q	q′	Rep.	Tr.	Rows	Cols.
SS I 1	2	2	1	1	2	2	5	5	25	4	10	10
SS I 2	2	4	1	1	1	1	6	6	18	8	12	12
SS I 3	3	3	1	2	2	0	2	7	14	9	6	21
SS I 4	3	3	2	2	0	0	2	2	4	9	6	6
SS I 5	3	3	2	2	1	0	2	5	10	9	6	15
SS I 6	3	6	2	2	0	0	4	4	8	18	12	12
SS I 7	4	2	1	2	2	0	1	3	9	8	4	18
SS I 8	4	2	2	2	1	1	2	2	18	8	12	12
SS I 9	4	2	3	2	0	0	1	1	3	8	4	6
SS I 10	4	2	3	2	0	1	3	1	9	8	6	12
SS I 11	4	2	3	2	0	2	5	1	15	8	6	20
SS I 12	4	2	3	2	3	0	1	5	15	8	4	30
SS I 13	4	3	2	2	1	0	1	3	9	12	6	18
SS I 14	4	4	3	2	0	0	2	2	6	16	8	12
SS I 15	4	4	3	3	0	0	3	3	9	16	12	12
SS I 16	6	3	5	2	0	0	1	1	5	18	6	15

TABLE III

Useful designs in Family HS I

Design	m	n	c	λ_2	k	p	p′	q	q′	Rep.	Tr.	Rows	Cols.
HS I 1	2	3	1	1	1	0	1	1	1	3	6	2	9
HS I 2	2	3	1	1	1	0	4	3	1	9	6	2	27
HS I 3	2	3	1	1	1	1	1	1	4	12	6	8	9
HS I 4	2	3	1	1	1	2	1	1	7	21	6	9	14
HS I 5	2	3	2	4	1	0	1	1	2	6	6	4	9
HS I 6	2	3	2	4	1	0	4	3	2	18	6	4	27
HS I 7	2	3	2	4	1	2	1	1	8	24	6	9	16
HS I 8	2	4	2	3	1	0	1	1	2	6	8	4	12
HS I 9	2	4	2	3	1	1	1	1	6	18	8	12	12
HS I 10	2	5	1	1	1	0	2	1	1	5	10	2	25
HS I 11	3	2	1	1	2	0	4	7	1	14	6	3	28
HS I 12	3	2	1	1	2	1	0	1	3	6	6	4	9
HS I 13	3	2	1	1	2	3	0	1	7	14	6	4	21
HS I 14	3	2	1	1	2	4	0	1	9	18	6	4	27
HS I 15	3	2	1	2	1	0	1	1	1	4	6	3	8
HS I 16	3	2	1	2	1	1	1	1	3	12	6	8	9

TABLE III—*Continued*

Design	m	n	c	λ_2	k	p	p′	q	q′	Rep.	Tr.	Rows	Cols.
HS I 17	3	2	1	2	2	0	2	2	1	8	6	3	16
HS I 18	3	2	1	2	2	1	2	2	3	24	6	9	16
HS I 19	3	4	1	1	1	0	1	1	1	4	12	3	16
HS I 20	3	4	3	9	1	0	1	1	3	12	12	9	16
HS I 21	3	5	1	1	2	0	1	1	1	5	15	3	25
HS I 22	4	2	1	3	2	1	1	1	2	18	8	12	12
HS I 23	4	3	1	1	1	0	2	3	1	9	12	4	27
HS I 24	4	3	1	1	3	0	0	1	1	3	12	4	9
HS I 25	4	3	1	1	3	1	0	1	4	12	12	9	16
HS I 26	4	3	2	4	3	0	0	1	2	6	12	8	9
HS I 27	4	5	1	1	1	0	1	1	1	5	20	4	25
HS I 28	5	2	1	2	4	0	0	1	1	4	10	5	8
HS I 29	5	2	1	2	4	1	0	1	3	12	10	8	15
HS I 30	5	2	1	3	1	0	1	1	1	6	10	5	12
HS I 31	5	2	1	7	4	0	2	1	1	14	10	5	28
HS I 32	5	3	1	2	1	0	1	1	1	6	15	5	18
HS I 33	5	3	1	3	4	0	1	1	1	9	15	5	27
HS I 34	5	4	1	1	4	0	0	1	1	4	20	5	16
HS I 35	6	5	1	1	5	0	0	1	1	5	30	6	25
HS I 36	7	2	1	3	6	0	0	1	1	6	14	7	12
HS I 37	7	2	1	4	1	0	1	1	1	8	14	7	16
HS I 38	7	3	1	2	6	0	0	1	1	6	21	7	18
HS I 39	9	2	1	4	8	0	0	1	1	8	18	9	16

TABLE IV

Useful designs in Family RS I

Design	m	n	k	k′	R	Λ_1	Λ_2	p	p′	q	q′	Rep.	Tr.	Rows	Cols.
RS I 1	2	2	2	1	5	1	2	0	2	1	1	5	4	2	10
RS I 2	2	2	2	1	5	3	1	0	2	1	1	5	4	2	10
RS I 3	2	2	2	1	5	1	2	2	2	1	5	25	4	10	10
RS I 4	2	2	2	1	5	3	1	2	2	1	5	25	4	10	10
RS I 5	2	2	2	1	7	1	3	0	3	1	1	7	4	2	14
RS I 6	2	2	2	1	7	3	2	0	3	1	1	7	4	2	14
RS I 7	2	2	2	1	7	5	1	0	3	1	1	7	4	2	14
RS I 8	2	2	2	1	11	1	5	0	5	1	1	11	4	2	22
RS I 9	2	2	2	1	11	3	4	0	5	1	1	11	4	2	22
RS I 10	2	2	2	1	11	5	3	0	5	1	1	11	4	2	22
RS I 11	2	2	2	1	11	7	2	0	5	1	1	11	4	2	22

TABLE IV—*Continued*

Design	m	n	k	k′	R	Λ_1	Λ_2	p	p′	q	q′	Rep.	Tr.	Rows	Cols.
RS I 12	2	2	2	1	11	9	1	0	5	1	1	11	4	2	22
RS I 13	2	2	2	1	13	1	6	0	6	1	1	13	4	2	26
RS I 14	2	2	2	1	13	3	5	0	6	1	1	13	4	2	26
RS I 15	2	2	2	1	13	5	4	0	6	1	1	13	4	2	26
RS I 16	2	2	2	1	13	7	3	0	6	1	1	13	4	2	26
RS I 17	2	2	2	1	13	9	2	0	6	1	1	13	4	2	26
RS I 18	2	2	2	1	13	11	1	0	6	1	1	13	4	2	26
RS I 19	2	3	2	1	7	2	1	0	3	1	1	7	6	2	21
RS I 20	2	3	2	1	9	3	1	0	4	1	1	9	6	2	27
RS I 21	2	3	4	1	14	9	8	0	3	1	2	14	6	4	21
RS I 22	2	3	4	1	18	12	10	0	4	1	2	18	6	4	27
RS I 23	2	4	4	1	10	6	3	0	2	1	2	10	8	4	20
RS I 24	2	5	4	1	10	5	2	0	2	1	2	10	10	4	25
RS I 25	3	2	3	1	7	2	3	0	2	1	1	7	6	3	14
RS I 26	3	2	3	1	7	2	3	1	2	1	3	21	6	9	14
RS I 27	3	2	3	1	13	2	6	0	4	1	1	13	6	3	26
RS I 28	3	2	3	1	13	6	5	0	4	1	1	13	6	3	26
RS I 29	3	2	3	2	7	2	3	0	4	2	1	14	6	3	28
RS I 30	3	2	3	2	8	4	3	0	2	1	1	8	6	3	16
RS I 31	3	2	3	2	8	4	3	1	2	1	3	24	6	9	16
RS I 32	3	2	3	2	11	2	5	0	3	1	1	11	6	3	22
RS I 33	3	2	3	2	11	6	4	0	3	1	1	11	6	3	22
RS I 34	3	2	3	2	14	8	5	0	4	1	1	14	6	3	28
RS I 35	3	3	3	1	7	1	2	0	2	1	1	7	9	3	21
RS I 36	3	3	3	1	7	4	1	0	2	1	1	7	9	3	21
RS I 37	3	3	3	1	10	1	3	0	3	1	1	10	9	3	30
RS I 38	3	3	3	1	10	4	2	0	3	1	1	10	9	3	30
RS I 39	3	3	3	1	10	7	1	0	3	1	1	10	9	3	30
RS I 40	3	3	3	2	5	2	1	0	1	1	1	5	9	3	15
RS I 41	3	3	6	1	14	8	9	0	2	1	2	14	9	6	21
RS I 42	3	3	6	1	14	11	8	0	2	1	2	14	9	6	21
RS I 43	3	3	6	2	10	7	6	0	1	1	2	10	9	6	15
RS I 44	3	4	3	1	7	2	1	0	2	1	1	7	12	3	28
RS I 45	4	2	4	1	9	3	4	0	2	1	1	9	8	4	18
RS I 46	4	2	4	1	13	3	6	0	3	1	1	13	8	4	26
RS I 47	4	2	4	1	13	9	5	0	3	1	1	13	8	4	26
RS I 48	4	2	4	3	11	3	5	0	2	1	1	11	8	4	22
RS I 49	4	2	4	3	11	9	4	0	2	1	1	11	8	4	22
RS I 50	4	2	4	3	15	3	7	0	3	1	1	15	8	4	30
RS I 51	4	2	4	3	15	9	6	0	3	1	1	15	8	4	30
RS I 52	4	4	4	3	7	3	1	0	1	1	1	7	16	4	28

TABLE V

Other useful group divisible designs

Design	s	t	m	n	k	k′	c	λ_2	c′	λ'_2	R	Λ_1	Λ_2	R′	Λ'_1	Λ'_2	u	h	r	p	p′	q	q′	Rep.	Tr.	Rows	Cols.
SS VI 1	3	–	9	2	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	0	0	1	1	8	18	12	12
SS VII 1	–	2	6	2	5	–	–	–	–	–	–	–	–	–	–	–	–	–	–	0	0	1	1	5	12	6	10
SS VII 2	–	3	8	2	7	–	–	–	–	–	–	–	–	–	–	–	–	–	–	0	0	1	1	7	16	8	14
HH 1	–	–	2	4	–	–	2	3	2	3	–	–	–	–	–	–	–	–	–	1	1	1	1	18	8	12	12
HH 2	–	–	4	2	–	–	1	3	1	3	–	–	–	–	–	–	–	–	–	1	1	1	1	18	8	12	12
HS V 1	–	1	7	2	–	–	1	5	–	–	–	–	–	–	–	–	–	–	–	0	1	1	1	10	14	7	20
HS VI 1	–	1	7	2	–	–	1	2	–	–	–	–	–	–	–	–	–	–	–	0	0	1	1	4	14	7	8
HS VI 2	–	2	11	2	–	–	1	3	–	–	–	–	–	–	–	–	–	–	–	0	0	1	1	6	22	11	12
RR I 1	–	–	2	2	2	2	–	–	–	–	5	1	2	5	1	2	–	–	–	2	2	1	1	25	4	10	10
RR I 2	–	–	2	2	2	2	–	–	–	–	5	1	2	5	3	1	–	–	–	2	2	1	1	25	4	10	10
RR I 3	–	–	2	2	2	2	–	–	–	–	5	3	1	5	3	1	–	–	–	2	2	1	1	25	4	10	10
RH I 1	–	–	2	3	3	–	1	1	–	–	7	4	2	–	–	–	–	–	–	1	2	1	1	21	6	9	14
RH I 2	–	–	2	3	3	–	2	4	–	–	8	5	2	–	–	–	–	–	–	1	2	1	1	24	6	9	16
RH I 3	–	–	3	2	4	–	1	1	–	–	14	10	8	–	–	–	–	–	–	0	3	1	1	14	6	4	21
RH I 4	–	–	3	2	4	–	1	1	–	–	18	10	11	–	–	–	–	–	–	0	4	1	1	18	6	4	27
RH I 5	–	–	3	2	4	–	1	1	–	–	18	14	10	–	–	–	–	–	–	0	4	1	1	18	6	4	27
RS VII 1	–	–	2	4	1	–	–	–	–	–	–	–	–	–	–	–	1	2	–	0	1	1	3	9	8	6	12
RS XIII 1	–	–	4	2	1	–	–	–	–	–	–	–	–	–	–	–	–	–	–	0	1	1	1	5	8	4	10
RS XIII 2	–	–	4	2	1	–	–	–	–	–	–	–	–	–	–	–	–	–	–	1	1	1	3	15	8	10	12
RS XVII 1	–	–	2	4	2	1	–	–	–	–	–	–	–	–	–	–	–	–	3	1	1	1	5	15	8	10	12

one treatment, if $n = 3$ the design is one in balanced incomplete blocks and if $n = 4$ it may be regarded as a group divisible design with 3 groups of 2 treatments each; hence for practical purposes $n > 4$. Further, in addition to the possibilities of extension by adding p complete replicates to each block and repeating the whole design, a times say, complete replicates of balanced incomplete block designs may be added. Thus, in fully extended form, the basic triangular design is as follows:

$$\text{T} \qquad \frac{q[2k + pn(n-1)]}{2k}, \frac{n(n-1)}{2}, \frac{qn(n-1)}{2k}, \frac{2k + pn(n-1)}{2},$$

where $q = ar + AR$,

$$\lambda = \frac{2R(k-1)}{(n-1)(n-2)}, n > 4$$

Thus the triangular O:PP design has the following form:

$$\text{TT} \qquad \frac{q[2k + pn(n-1)}{2k}, \frac{n(n-1)}{2}, \frac{qn(n-1)}{2k}, \frac{q'n(n-1)}{2k'},$$

where

$$q = \frac{k[2k' + p'n(n-1)]}{n(n-1)}, n > 4,$$

$$q' = \frac{k'[2k + pn(n-1)]}{n(n-1)}$$

and there are the further restrictions above on q and similarly on q'.

TABLE VI

Smallest possible designs in families with no useful members

Family	s	w	t	m	n	k	k′	c	λ_2	R	Λ_1	Λ_2	u	h	u′	h′	r	r′	p	p′	q	q′	Rep.	Tr.	Rows	Cols.
SS II	3	–	–	9	3	8	–	–	–	–	–	–	–	–	–	–	–	–	0	0	2	1	8	27	9	24
SS III	3	–	–	9	2	–	–	–	–	–	–	–	–	–	–	–	–	–	1	1	2	2	32	18	24	24
SS IV	3	–	–	9	3	8	–	–	–	–	–	–	–	–	–	–	–	–	0	0	2	2	16	27	18	24
SS V	3	–	–	9	2	–	–	–	–	–	–	–	–	–	–	–	–	–	0	1	2	1	16	18	12	24
SS VIII	–	2	–	16	3	–	–	–	–	–	–	–	–	–	–	–	–	–	2	1	2	6	150	48	60	120
SS IX	–	2	–	16	5	–	–	–	–	–	–	–	–	–	–	–	–	–	0	0	2	2	30	80	40	60
SS X	–	–	2	6	2	–	–	–	–	–	–	–	–	–	–	–	–	–	2	2	3	3	75	12	30	30
SS XI	–	–	1	7	7	1	–	–	–	–	–	–	–	–	–	–	–	–	1	0	1	10	10	49	7	70
SS XII	–	–	1	7	7	–	–	–	–	–	–	–	–	–	–	–	–	–	0	0	3	3	9	49	21	21
SS XIII	–	–	1	7	7	1	–	–	–	–	–	–	–	–	–	–	–	–	1	0	1	11	11	49	7	77
SS XIV	–	–	1	7	7	–	–	–	–	–	–	–	–	–	–	–	–	–	0	0	3	4	12	49	21	28
SS XV	–	–	1	7	7	–	–	–	–	–	–	–	–	–	–	–	–	–	0	0	4	4	16	49	28	28
HS II	3	–	–	9	3	–	–	1	4	–	–	–	–	–	–	–	–	–	1	1	1	3	48	27	36	36
HS III	3	–	–	9	3	–	–	1	5	–	–	–	–	–	–	–	–	–	1	1	1	3	60	27	36	45
HS IV	–	–	2	6	3	–	–	2	12	–	–	–	–	–	–	–	–	–	1	1	1	3	45	18	27	30
RR II	–	–	–	2	4	4	–	–	–	11	7	3	1	2	–	–	–	–	2	1	1	1	33	8	12	22
RR III	–	–	–	2	9	–	–	–	–	–	–	–	1	3	1	3	–	–	18	18	2	2	6272	18	336	336
RR IV	–	–	–	4	2	2	–	–	–	9	3	1	–	–	–	–	–	–	4	1	1	1	45	8	10	36
RR V	–	–	–	4	2	–	–	–	–	–	–	–	–	–	–	–	–	–	2	2	2	2	50	8	20	20
RR VI	–	–	–	2	4	2	4	–	–	9	5	3	–	–	–	–	3	–	2	1	1	1	27	8	12	18
RR VII	–	–	–	2	6	4	–	–	–	–	–	–	1	3	–	–	5	–	3	3	3	1	150	12	40	45
RR VIII	–	–	–	2	6	4	3	–	–	–	–	–	–	–	–	–	5	5	3	1	1	2	50	12	15	40
RH II	–	–	–	2	6	–	–	3	5	–	–	–	1	2	–	–	–	–	1	2	1	1	50	12	20	30
RH III	–	–	–	4	2	–	–	1	3	–	–	–	–	–	–	–	–	–	1	2	2	1	30	8	12	20
RH IV	–	–	–	3	6	2	–	3	5	–	–	–	–	–	–	–	5	–	1	2	1	1	50	18	20	45
RS II	3	–	–	9	3	9	–	–	–	22	4	7	–	–	–	–	–	–	1	7	3	3	264	27	36	198
RS III	3	–	–	9	3	9	–	–	–	17	8	5	–	–	–	–	–	–	1	5	3	3	204	27	36	153
RS IV	–	–	2	6	2	6	–	–	–	9	5	4	–	–	–	–	–	–	2	1	1	3	45	12	18	30
RS V	–	–	1	7	2	7	–	–	–	17	6	8	–	–	–	–	–	–	0	2	1	1	17	14	7	34
RS VI	–	–	1	7	2	7	–	–	–	25	6	12	–	–	–	–	–	–	0	3	1	1	25	14	7	50
RS VIII	3	–	–	9	9	–	–	–	–	–	–	–	8	3	–	–	–	–	1	9	1	13	1456	81	156	756
RS IX	3	–	–	9	9	–	–	–	–	–	–	–	8	3	–	–	–	–	1	18	2	13	2912	81	156	1512
RS X	–	–	2	6	4	–	–	–	–	–	–	–	5	2	–	–	–	–	2	1	1	7	105	24	36	70
RS XI	–	–	1	7	14	–	–	–	–	–	–	–	6	7	–	–	–	–	0	735	3	13	66924	98	91	72072
RS XII	–	–	1	7	14	–	–	–	–	–	–	–	6	7	–	–	–	–	0	980	4	13	89232	98	91	96096
RS XIV	–	2	–	16	8	–	–	–	–	–	–	–	–	–	–	–	–	–	3	17	12	20	6900	128	400	2208
RS XV	–	2	–	16	8	–	–	–	–	–	–	–	–	–	–	–	–	–	3	5	4	20	2300	128	400	736
RS XVI	–	–	3	8	4	–	–	–	–	–	–	–	–	–	–	–	–	–	5	5	4	12	924	32	168	176
RS XVIII	3	–	–	9	5	3	–	–	–	–	–	–	–	–	–	–	2	–	1	1	2	4	64	45	48	60
RS XIX	3	–	–	9	5	3	–	–	–	–	–	–	–	–	–	–	2	–	1	0	1	4	32	45	30	48
RS XX	–	–	2	6	4	2	–	–	–	–	–	–	–	–	–	–	3	–	2	1	1	5	75	24	36	50
RS XXI	–	–	1	7	10	7	–	–	–	–	–	–	–	–	–	–	3	–	1	0	1	11	33	70	30	77
RS XXII	–	–	1	7	9	7	–	–	–	–	–	–	–	–	–	–	2	–	1	0	2	10	40	63	36	70

TABLE VII

Useful designs in family TT

Design	n	k	k′	r	a	R	A	r′	a′	R′	A′	p	p′	q	q′	Rep.	Tr.	Rows	Cols.
TT 1	5	4	5	2	1	0	0	3	1	9	1	2	0	2	12	12	10	5	24
TT 2	5	4	5	2	1	0	0	3	4	0	0	2	0	2	12	12	10	5	24
TT 3	5	5	6	3	1	0	0	3	1	0	0	0	0	3	3	3	10	5	6
TT 4	5	5	6	3	1	0	0	3	2	9	1	2	0	3	15	15	10	6	25
TT 5	5	5	6	3	1	0	0	3	5	0	0	2	0	3	15	15	10	6	25
TT 6	5	5	6	3	1	0	0	6	1	9	1	2	0	3	15	15	10	6	25
TT 7	6	5	6	2	1	0	0	4	2	0	0	1	0	2	8	8	15	6	20
TT 8	6	6	10	4	1	0	0	4	1	0	0	0	0	4	4	4	15	6	10

The useful designs in family TT are given in Table VII, where it is necessary to include the auxiliary parameters r, a, R, A, r', a', R', A', to differentiate between designs otherwise identical. It will be seen that in all of them the basic triangular designs are singly linked blocks (SLB), $r = 2$, $k = n - 1$, or their complement, $r = n - 2$, $k = (n - 1)(n - 2)/2$, one way and doubly linked blocks (DLB), $r = n - 2$, $k = n$, the other way.

These values of r and k give rise to the following triangular designs:

SLB $$\frac{q(pn + 2)}{2}, \frac{n(n - 1)}{2}, \frac{qn}{2}, \frac{(pn + 2)(n - 1)}{2}$$

Complement of SLB $$\frac{q(pn + n - 2)}{n - 2}, \frac{n(n - 1)}{2}, \frac{qn}{n - 2}, \frac{(pn + n - 2)(n - 1)}{2}$$

DLB $$\frac{q(pn - p + 2)}{2}, \frac{n(n - 1)}{2}, \frac{q(n - 1)}{2}, \frac{n(pn - p + 2)}{2}$$

Thus, for an O:PP design to be SLB both ways or SLB one way and its complement the other $qn = (p'n + 2)(n - 1)$ or $qn = (p'n + n - 2)(n - 1)$ respectively; it is easily seen that neither of these equations has any integral solutions for $n > 2$. If an O:PP design is the complement of SLB both ways then $2qn = (p'n + n - 2)(n - 1)(n - 2)$. n has no factor in common with $(n - 1)$ nor, if odd, with $(n - 2)$; thus $n/2$ is integral, $= x$ say, and $qx = (p'x + x - 1)(2x - 1)(x - 1)$, which is impossible for $x > 1$, i.e., for $n > 2$. For an O:PP design to be DLB both ways $q(n - 1) = n(p'n - p' + 2)$, which is impossible for $n > 3$.

Thus, no O:PP designs can be SLB or its complement both ways, neither can they be DLB both ways, but there is no reason why designs of either kind should not fit with other triangular designs to make an O:PP design.

Latin square designs. The basic Latin square design has r replicates of n^2 treatments on rn^2/k blocks of k plots each but, as for the triangular design, it can be extended by adding on complete replicates to each block, repeating the whole design and adding on balanced incomplete blocks. Thus, in fully extended form, the basic Latin square design is:

L $$\frac{q(k + pn^2)}{k}, n^2, \frac{qn^2}{k}, k + pn^2, \text{ where } q = ar + AR, \lambda = \frac{R(k - 1)}{n^2 - 1}$$

Thus the Latin square O:PP design has the following parameters:

LL $$\frac{q(k + pn^2)}{k}, n^2, \frac{qn^2}{k}, \frac{q'n^2}{k'}, \text{ where } q = \frac{k(k' + p'n^2)}{n^2}, q' = \frac{k'(k + pn^2)}{n^2}$$

and there are the further restrictions above on q and similarly on q'.

There are only two useful designs in the family LL. Both have $n = 3$, $k = k' = 6$, $p = p' = 0$ and so $q = q' = 4$ while $a = a' = 1$, $r = r' = 4$, $A = A' = R = R' = 0$, thus leading to designs with 4 replicates of 9 treatments on 6 rows and columns. The only difference between the designs is that in one

first associates concur three times and second associates twice in both rows and columns, while in the other first associates concur three times in rows and twice in columns and conversely for second associates. Thus, if the designs are LL 1 and LL 2 respectively, then, using the notation previously adopted (Freeman, 1957b), in LL 1

$$\lambda_1 = \mu_1 = 3, \lambda_2 = \mu_2 = 2 \text{ and in LL2 } \lambda_1 = \mu_2 = 3, \lambda_2 = \mu_1 = 2.$$

Since, for LL 2, $\nu_1 = \nu_2 = 30$, in the same notation, the design has equal efficiency for both types of associates, and it is the only useful O:PP design with this property.

Summary. All known families of O:PP designs with two associate-classes are classified, these including all with at least one member of practicable size, i.e., having more than two replicates or treatments, not more than 30 replicates, treatments, rows or columns, and not more than 150 plots in all. The designs within these limits are tabulated in their families and where a family has no practicable design its smallest member is given.

Acknowledgements. I should like to express my thanks to Dr. N. L. Johnson of University College, London, for his advice throughout the preparation of this paper, particularly with regard to the tabulation of the designs.

REFERENCES

J. W. Archbold, and N. L. Johnson, "A method of constructing partially balanced incomplete block designs," *Ann. Math. Stat.*, 27, (1956), pp. 624-32.

R. C. Bose, "On the construction of balanced incomplete block designs," *Ann. Eugen.*, 9, (1939), pp. 353-99.

R. C. Bose, W. H. Clatworthy, and S. S. Shrikhande, "Tables of partially balanced designs with two associate classes," *N. Carolina Agric. Expt. Sta. Tech. Bull.* No. 107 (1954).

R. C. Bose, and W. S. Connor, "Combinatorial properties of group divisible incomplete block designs," *Ann. Math. Stat.*, 23, (1952), pp. 367-83.

R. C. Bose, S. S. Shrikhande, and K. N. Bhattacharya, "On the construction of group divisible incomplete block designs," *Ann. Math. Stat.*, 24, (1953), pp. 167-95.

G. H. Freeman, "Some further methods of constructing regular group divisible incomplete block designs," *Ann. Math. Stat.* 28, (1957a), pp. 479-87.

G. H. Freeman, "Some experimental designs of use in changing from one set of treatments to another," Part I. *J. R. Statist. Soc.*, B 19, (1957b), pp. 154-62.

G. H. Freeman, "Some experimental designs of use in changing from one set of treatments to another," Part II. Existence of the designs. *J. R. Statist. Soc.*, B 19, (1957c), pp. 163-5.

T. N. Hoblyn, S. C. Pearce, and G. H. Freeman, "Some considerations in the design of successive experiments in fruit plantations," *Biometrics*, 10, (1954), pp. 503-15.

MOST ECONOMICAL MULTIPLE-DECISION RULES[1]

By Wm. Jackson Hall

University of North Carolina

0. Summary. This paper is concerned with non-sequential multiple-decision procedures for which the sample size is a minimum subject to either (1) lower bounds on the probabilities of making correct decisions or (2) upper bounds on the probabilities of making incorrect decisions. Such decision procedures are obtained by constructing artificial decision problems for which the minimax strategies provide solutions to problems (1) and (2). These are shown to be "likelihood ratio" and "unlikelihood ratio" decision rules, respectively. Thus, although problems (1) and (2) are formulated in the spirit of the classical Neyman-Pearson approach to two-decision problems, minimax theory is used as a tool for their solution.

Problems of both "simple" and "composite" discrimination are considered and some examples indicated. (Some multivariate examples are given in [4].) Various properties of the decision rules are derived, and relationships with works of Wald, Lindley, Rao and others are cited.

1. Simple discrimination.

A. *Formulation of the problem.* We are concerned with a sequence $X_1, X_2, \cdots$, of real- or vector-valued, independent, and identically distributed random variables, each having a density function f, belonging to some specified class Ω, w.r.t. a fixed measure μ.

The decision problem is to formulate a rule for choosing a non-negative integer n (completely non-random), and, after taking an observation

$$x = (x_1, \cdots, x_n)$$

on $X = (X_1, \cdots, X_n)$, for choosing one of m possible alternative decisions $A_1, \cdots, A_m$. A multiple decision rule (m-d.r.) for choosing among $A_1, \cdots, A_m$ on the basis of x is defined by an ordered set of non-negative, real-valued, measurable functions $\phi(x) = [\phi_1(x), \cdots, \phi_m(x)]$ on the space $\mathfrak{X}$ of x such that $\sum_i \phi_i = 1$ identically in x (for $n = 0$, the ϕ_i's are constants). A_i is then chosen with probability $\phi_i(x)$ when x is observed. For non-randomized d.r.'s (all ϕ_i's equal 0 or 1), the ϕ_i's are characteristic functions of mutually exclusive and exhaustive "acceptance" regions $R_1, \cdots, R_m$ in $\mathfrak{X}$, where A_i is accepted if $x \in R_i$.

Received January 24, 1956; revised November 15, 1957.

[1] This research was sponsored partly by the Office of Naval Research under Contract No. Nonr-855(06) for research in probability and statistics at Chapel Hill and partly by the United States Air Force through the Air Force Office of Scientific Research of the Air Research and Development Command, under Contract No. 18(600)-458. Reproduction in whole or in part for any purpose of the United States Government is permitted.

A subscript or superscript n denotes the corresponding sample size; $f^n(x)$ and μ^n are the joint density and product measure, respectively.

We suppose throughout Section 1 that Ω consists of a finite number, say l, of elements $f_1, \cdots, f_l$; we say that the corresponding decision problem is one of "simple discrimination" and a d.r. is a d.r. for "discriminating among $f_1, \cdots, f_l$." Here, if μ is non-atomic, only non-randomized d.r.'s need be considered [2].

A d.r. $D = D_n$ is characterized by the functions

$$p_{ij}(D) = \Pr(D \text{ chooses } A_j \mid f_i) = \int_{\mathfrak{X}} \phi_j(x) f_i^n(x)\, d\mu^n \quad (i = 1, \cdots, l; j = 1, \cdots, m).$$

We consider two different criteria for choosing a d.r. for simple discrimination. The first assumes that $l = m$ and that the decision A_i is to be preferred when f_i is true. Denote $p_{ii}(D) = p_i(D) = 1 - q_i(D)$, so that p_i is the probability of a "correct" decision and q_i the probability of an "incorrect" decision when f_i is true.

Definition 1. Let $\alpha = (\alpha_1, \cdots, \alpha_m)$ be a given vector of positive constants each less than one. A d.r. D_N, based on a sample of size N, is said to be a most economical m-decision rule relative to the vector α for discriminating among $f_1, \cdots, f_m$ if it satisfies

$$p_i(D) \geqq \alpha_i \qquad (i = 1, \cdots, m) \tag{1}$$

and if N is the least integer n for which (1) may be satisfied by some m-d.r. D_n based on a sample of size n. N is said to be the most economical sample size.

We now no longer require that $l = m$, but suppose that corresponding to each f_i one or more of the alternatives A_j is preferable, or "correct," when f_i is true.

Definition 2. Let $\beta = (\beta_{ij})$ be a given $l \times m$ matrix of positive constants such that for every i, j pair for which A_j is a correct decision when f_i is true $\beta_{ij} = 1$. A d.r. D_N, based on a sample of size N, is said to be a most economical m-decision rule relative to the matrix β for discriminating among $f_1, \cdots, f_l$ if it satisfies

$$p_{ij}(D) \leqq \beta_{ij} \quad (i = 1, \cdots, l; j = 1, \cdots, m) \tag{2}$$

and if N is the least integer n for which (2) may be satisfied by some m-d.r. D_n based on a sample of size n. N is said to be the most economical sample size.

If $l = m$ and A_i is preferred when f_i is true, then an M.E. d.r. relative to β also controls the probabilities of correct decisions if $\sum_{j \neq i} \beta_{ij} < 1$ for all i.

If $l = m = 2$, both (1) and (2) reduce to upper bounds on the probabilities of the two kinds of error, and Definitions 1 and 2 define an M.E. 2-d.r. as one with minimum sample size subject to these bounds.

It is intuitively clear (and elementary to prove) that a necessary and sufficient condition for the existence of a M.E. m-d.r. relative to any α or β ($l = m$) is that there exist uniformly consistent sequences of 2-d.r.'s for discriminating between every pair $\omega_i, \omega_j (i \neq j)$ [5].

We shall utilize elements of Wald's theory of decision functions as given in

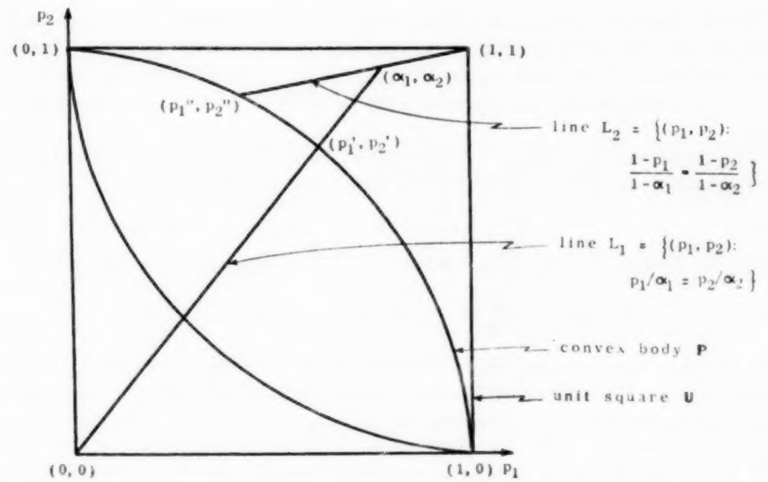


FIG. 1

[14], and shall use in particular some of the results of Sections 3.5 and 5.1.1, altering his notation slightly. The differences in the "data of the decision problem" assumed by Wald and here are only minor.

Let $\mathfrak{D}_n$ denote the class of all m-d.r.'s based on a sample of size n ($n = 0, 1, 2, \cdots$). Clearly, for $n \leqq N$, $\mathfrak{D}_n \subseteq \mathfrak{D}_N$; Lemma 1 follows almost immediately.

LEMMA 1. *For every fixed sample size* $n = 0, 1, 2, \cdots$, *let* D_n^0 *be a minimax d.r. and denote* $r_n = \max_i r(f_i, D_n^0)$, *where* $r(f_i, D_n)$ *is the risk w.r.t. some bounded loss function. Then the sequence* $\{r_n\}$, $n = 0, 1, 2, \cdots$, *is a non-increasing sequence.*

B. *Most economical decision rules relative to a vector* α. Later in this section, we shall apply Wald's theory to two specific loss functions and develop in each case a method of obtaining M.E. d.r.'s as defined by Definition 1. First, we motivate geometrically the selection of such loss functions so as to identify the minimax strategy with the desired one. This alternative approach may give some geometrical insight into the properties of the d.r.'s obtained.[2]

For fixed n, let $p(D) = (p_1(D), \cdots, p_m(D))$ denote a point in m-space, and $P_n = \{p(D): D \in \mathfrak{D}_n\}$. It can be shown ([2], [10]) that P_n is a convex body in the unit m-cube U containing all corners of U with coordinates summing to unity. The case $m = 2$ is illustrated. (Conditions under which P_n is a proper subset of $P_{n'}$ for $n < n'$ and for which P_n tends with increasing n to U are given elsewhere ([3], [5]).)

---

[2] The author is indebted to the referee for considerable improvement of this geometric presentation.

In the diagram, the point $\alpha \varepsilon P_n$; therefore n is smaller than the required sample size. The M.E. sample size is the smallest n for which $\alpha \varepsilon P_n$, in which case the points p', p'' and α coincide (approximately). To test whether or not $\alpha \varepsilon P_n$, we can examine the position of the points p' or p'' relative to the position of α.

The points on the "upper" surface of P_n (n fixed) include all points $p(D_n^*)$ corresponding to Bayes strategies D_n^* when the loss function is

$$W(f_i, A_j) = W_{ij} = -\delta_{ij}/\alpha_i \qquad (i, j = 1, \cdots, m), \tag{3}$$

where δ_{ij} denotes the Kronecker δ-function.[3] Then the risk w.r.t. W_{ij} is

$$r(f_i, D) = -\sum_{j=1}^{m} \delta_{ij}\, p_{ij}(D)/\alpha_i = -p_i(D)/\alpha_i \qquad (i = 1, \cdots, m). \tag{4}$$

If p' is not on the boundary of U, the least favorable distribution for the weight function W_{ij} will positively weigh each element in Ω. (This will occur if the region in $\mathfrak{X}$ of positive density is constant over Ω.) In this case, the minimax strategy D_n' will be such that $p(D_n')$ is on the line L_1 of constant risk; i.e., $p(D_n') = p'$. (To obtain the minimax geometry with the loss function W_{ij}, transform the diagram by dividing the ith coordinate by $-\alpha_i$; then the convex body P_n goes into the convex body of risk points $(r_1, \cdots, r_m)$, $r_i = r(f_i, D)$.)

Alternatively, the "upper" surface of P_n corresponds to Bayes strategies when the loss function is

$$W^*(f_i, A_j) = W_{ij}^* = (1 - \delta_{ij})/(1 - \alpha_j) \qquad (i, j = 1, 2, \cdots, m) \tag{5}$$

and the risk function is $r^*(f_i, D) = q_i(D)/\beta_i$, where $\beta_i = 1 - \alpha_i$. The least favorable distribution will likewise positively weight each element of Ω whenever p'' is not on the boundary of U. In this case, the minimax strategy D_n'' will be such that $p(D_n'')$ is on the line L_2 of constant risk; i.e., $p(D_n'') = p''$. (To obtain the minimax geometry with W_{ij}^*, transform the diagram by dividing the ith coordinate by $1 - \alpha_i$, again transforming P_n into the convex body of risk points.) This latter approach is similar to that used by Rao [11] for problems of classification in multivariate analysis.

When $l = m > 2$, there is an added complication for the latter loss function since the line (L_2) from $(1, 1, \cdots, 1)$ through α need not necessarily pierce P for $n < N$, the M.E. sample size. (Of course, if $\alpha \varepsilon P_n$, then the line certainly pierces P_n.) Thus the components of a least favorable distribution are not necessarily positive unless $n \geqq N$ and p'' is in the interior of U.

Thus, in one instance, minimax rules maximize the common ratio $p_1/\alpha_1 = \cdots = p_m/\alpha_m$ and, in the other, minimize the common ratio $q_1/\beta_1 = \cdots = q_m/\beta_m$. The M.E. sample size is the smallest one for which the common ratio is $\geqq 1$ or $\leqq 1$, respectively. We now formalize these results. (Wald's Theorem 5.3[4] asserts the existence of a minimax d.r. D^0 for any (fixed) sample size.)

[3] This loss function satisfies Wald's requirements although it is not necessarily zero when a correct decision is made nor necessarily positive otherwise, as intuitively suggested, but never required mathematically, by Wald.

[4] All references to Wald refer to [14] unless otherwise specified.

THEOREM 1. *For each* $n = 0, 1, 2, \cdots$, *let* D_n^0 *be a minimax d.r. w.r.t. the weight function* (3) *for samples of fixed size* n. *Suppose for some* n,

$$\max_i r(f_i, D_n^0) \leqq -1 \tag{6}$$

and let N *be the least such integer. Then* D_N^0 *is an M.E. d.r. relative to the vector* α *for discriminating among* $f_1, \cdots, f_m$. *Conversely, if there exists an M.E. d.r. relative to* α *for discriminating among* $f_1, \cdots, f_m$, *and the M.E. sample size is* N, *then* D_N^0 *is an M.E. d.r.*

PROOF. From (4) and (6), it follows that D_N^0 satisfies (1). Now suppose for some $n < N$, there exists a d.r. D_n satisfying (1). Since D_n^0 is minimax, $\max_i r(f_i, D_n^0) \leqq \max_i r(f_i, D_n) = \max_i [-p_i(D_n)/\alpha_i]$. Since D_n satisfies (1), we have from above that $\max_i r(f_i, D_n^0) \leqq -1$, in contradiction to the fact that N is the least integer n for which this is true. Hence, D_N^0 is an M.E. d.r.

To prove the converse, suppose D_N is an M.E. d.r. Then

$$-1 \geqq \max_i [-p_i(D_N)/\alpha_i] = \max_i r(f_i, D_N) \geqq \max_i r(f_i, D_N^0)$$

since D_N^0 is a minimax d.r. Hence, (6) is satisfied for $n = N$, and since N is the M.E. sample size, D_N^0 is an M.E. d.r.

Lemma 1 assures us that any n for which (6) is violated is too small. Now let us consider the structure of minimax d.r.'s for a fixed sample size n.

DEFINITION 3. A d.r. D defined by $\phi(x)$ is said to be a likelihood ratio d.r. if there exist positive constants $a_1, \cdots, a_m$ such that for any j and any x for which $\phi_j(x) > 0$, $a_j f_j^n(x) \geqq a_i f_i^n(x)$ for all $i \neq j$.
(Note that $a_1, \cdots, a_m$ determine ϕ completely except in sets of x for which $a_i f_i^n(x) = \max_j a_j f_j^n(x)$ for more than one value of i.) Setting $a_j = \xi_j/\alpha_j$, where $\xi = (\xi_1, \cdots, \xi_m)$ is an a priori distribution over $\Omega = (f_1, \cdots, f_m)$, it follows from Wald's Theorem 5.1 (with (5.6) replaced by (5.7)) that a Bayes d.r. relative to any ξ for which all $\xi_i > 0$ is a likelihood ratio d.r., and conversely.

Wald's Theorem 5.3 asserts the existence of a minimax d.r. and a least favorable distribution, and that any minimax d.r. is a Bayes d.r. relative to any least favorable distribution. Moreover, it follows from (4) and Wald's Theorem 5.3 (iii) that if all components of a least favorable distribution are positive, any minimax d.r. D^0 has the property:

$$p_1(D^0)/\alpha_1 = \cdots = p_m(D^0)/\alpha_m . \tag{7}$$

We shall give sufficient conditions for this to be true.

ASSUMPTION 1. If R is a subset of $\mathfrak{X}$ for which $\int_R f_i^n(x)\, d\mu^n = 0$ for some i, then $\int_R f_i^n(x)\, d\mu^n = 0$ for all values of i. (Whenever this assumption is made, we shall tacitly assume that $\mathfrak{X}$ is redefined so that $f_i^n(x) > 0$ for all i and $x \in \mathfrak{X}$.)

We state a theorem analogous to Wald's Theorem 5.4;[5] the proof (not given) is also analogous.

[5] It might be noted that Wald's condition (iii) of Theorem 5.4 is superfluous since it is always fulfilled; e.g., in Wald's notation, let $\delta = 1/u$ $(i = 1, \cdots, u)$ identically in x, and then $r(F_j, \delta) = (u-1)/u < 1$ for $j = 1, \cdots, k$.

THEOREM 2. *If Assumption* 1 *holds, all components of a least favorable distribution* ξ^0 *w.r.t. the weight function* w_{ij} *are positive.*

Hence, under Assumption 1, an M.E. d.r. may be obtained by the following method: for each sample size n, find a likelihood ratio d.r. D_n^0 for the constants $a_1, \cdots, a_m$ determined by Eqs. (7), and then choose N as the minimum n for which $p_1(D_n^0) \geqq \alpha_1$.

As an alternative approach, we can consider the weight function W_{ij}^* and proceed analogously to the first approach, giving a theorem identical to Theorem 1 with (6) replaced by $\max_i r(f_i, D_n^0) \leqq 1$; and, replacing $a_j = \xi_j/\alpha_j$ by ξ_j/β_j, it follows analogously that a Bayes d.r. relative to any ξ for which all $\xi_i > 0$ is a likelihood ratio d.r., and conversely. Moreover, if all components of a least favorable distribution are positive, any minimax d.r. D^0 has the property:

$$q_1(D^0)/\beta_1 = \cdots = q_m(D^0)/\beta_m. \tag{8}$$

We shall give sufficient conditions for this to be true. Analogously to Wald's Theorem 5.4, we have:

LEMMA 2. *If Assumption* 1 *holds, and if there exists some d.r. D for which*

$$r(f_i, D) < 1/\max_{1 \leqq j \leqq m} \beta_j \qquad (i = 1, \cdots, m),$$

then all components of a least favorable distribution are positive.

The following lemma may be useful in this regard:

LEMMA 3. *If* $\beta_i < [1/(m-1)]\sum_{j=1}^m \beta_j$ (*i.e.*, $\alpha_i > [\sum \alpha_j - 1]/[m-1]$) *for all i, then there exists a d.r. D for which* $r(f_i, D) < 1/\max_j \beta_j$ *for all i.*

The proof follows by considering a d.r. defined by $\phi_i(x) = 1 - (m-1)\beta_i/\sum \beta_j > 0$ identically in x $(i = 1, \cdots, m)$.

THEOREM 3. *Suppose Assumption* 1 *holds. For any sample size greater than or equal to the M.E. sample size, all components of a least favorable distribution are positive.*

PROOF. Suppose $n \geqq N$, the M.E. sample size, and that D_n^0 is a minimax d.r. for samples of size n; then, using Lemma 1 and the theorem analogous to Theorem 1, D_n^0 satisfies (1). Use of Lemma 2 completes the proof.

Hence, under Assumption 1, D_N^0 is a likelihood ratio d.r., and an M.E. d.r. may be obtained by considering likelihood ratio d.r.'s D_n^0 for each n for constants $a_1, \cdots, a_m$ determined by (8), and then choosing N as the minimum n for which $q_1(D_n^0) \leqq \beta_1$. If for some n one of the components of a least favorable distribution is zero, we know that n is less than the M.E. sample size (Lemma 1).

A Bayes d.r. relative to any ξ of which all components are positive is admissible [15]. Hence, any likelihood ratio d.r. is admissible, and under Assumption 1 M.E. d.r.'s obtained by either of the above approaches are admissible. Thus, denoting an M.E. d.r. by D_N^0, there does not exist a d.r. D_N' for which $p_i(D_N') \geqq p_i(D_N^0)$ $(i = 1, \cdots, m)$ with strict inequality for at least one i (under Assumption 1).

Suppose now that a real-valued statistic $t = t(x_1, \cdots, x_n)$ exists which is sufficient for the class $\{f_i^n\}$ $(i = 1, \cdots, m)$, and suppose that t has a monotone

likelihood ratio for some ordering of the elements of Ω; i.e., if $g_i(t)$ is the density of t corresponding to $f_i(x)$, then, for some ordering of the subscripts,

$$g_i(t_1)g_j(t_2) \geqq g_i(t_2)g_j(t_1)$$

for $i > j$ and $t_1 > t_2$ [8]. It follows almost immediately that for any $\phi(x)$ which defines a likelihood ratio d.r. there exist constants $\{c_i\}$, $-\infty = c_0 \leqq c_1 \leqq \cdots \leqq c_{m-1} \leqq c_m = \infty$, such that $\phi_i(x) > 0$ implies $c_{i-1} \leqq t(x) \leqq c_i$. Moreover, $\phi_i(x) = 1$ if the latter inequalities are strict, so that randomization may be required only at the points $t = c_i$ and only then if such points have positive probability. Such d.r.'s have been called monotone [1], [8]. If, for example, f_i is of the exponential type $f_i = \beta(\theta_i)e^{\theta_i x}r(x)$, $r \geqq 0$ and θ_i real, for all i, the above conditions are satisfied [1].

Example 1. Suppose f_i is a normal density function with mean θ_i $(-\infty < \theta_1 < \cdots < \theta_m < \infty)$ and known variance σ^2. Then $t = \bar{x}$ is sufficient and the c_i's and N may be obtained by first solving the following equations (iteratively) for the c_i^n's and n with $\rho_n = 1$:

$$(9) \qquad p_i(D_n) = \Phi[\sqrt{n}(c_i^n - \theta_i)/\sigma] - \Phi[\sqrt{n}(c_{i-1}^n - \theta_i)/\sigma] = \alpha_i\rho_n \qquad (i = 1, \cdots, n),$$

where Φ denotes the standard normal distribution function, and then, choosing N to be the least integer $\geqq n$, re-solving for the c_i's and ρ_N. Such a monotone rule will be minimax w.r.t. W_{ij} for the M.E. sample size. Alternatively, (9) may be replaced by equations of the form $1 - p_i(D_n) = (1 - \alpha_i)\rho_n'$, and a solution obtained which will be minimax w.r.t. W_{ij}^*.

Other examples may be treated analogously, allowing for randomization in the discrete cases if desired.

C. *Most economical decision rules relative to a matrix* β. To obtain M.E. d.r.'s as defined by Definition 2, we shall construct an artificial decision problem whose minimax solution will have the properties desired. For convenience, we replace each β_{ij} which is equal to unity by $+\infty$.

Suppose n fixed, and let Ω' be a set of density functions g_{ij} w.r.t. μ $(i = 1, \cdots, l; j = 1, \cdots, m)$, where $g_{ij} = f_i$ identically in x. Define a weight function $W(g_{ij}, A_k) = W_{ijk}$, where

$$(10) \quad W_{ijk} = 1/\beta_{ij} \quad \text{if } j = k \ (i = 1, \cdots, l; j, k = 1, \cdots, m) \text{ and } 0 \text{ otherwise.}$$

We consider the artificial decision problem of choosing among $A_1, \cdots, A_m$ when one of the $l' = lm$ density functions g_{ij} is "true", and where the "loss" incurred by choosing A_k when g_{ij} is "true", is $W(g_{ij}, A_k)$. The risk function is $r(g_{ij}, D) = \sum_k W_{ijk}p_{ijk}'(D)$, where $p_{ijk}'(D) = \Pr(D \text{ chooses } A_k \mid g_{ij}) = p_{ik}(D)$; thus $r(g_{ij}, D) = p_{ij}(D)/\beta_{ij}$ $(i = 1, \cdots, l;\ j = 1, \cdots, m)$. Wald's Theorem 5.3 asserts the existence of minimax d.r.'s.

THEOREM 4. *For each* $n = 0, 1, 2, \cdots$, *let* D_n^0 *be a minimax d.r. w.r.t. the weight function* (10) *for discriminating among* $g_{11}, g_{12}, \cdots, g_{lm}$ *for samples of fixed size* n. *Suppose for some* n, $\max_{i,j} r(g_{ij}, D_n^0) \leqq 1$, *and let* N *be the least such integer.*

Then D_N^0 is an M.E. d.r. relative to the matrix β for discriminating among $f_1, \cdots, f_l$. Conversely, if there exists an M.E. d.r. relative to β and N is the M.E. sample size, then D_N^0 is an M.E. d.r.
The theorem may be proved in a similar manner to Theorem 1. Now let us consider the structure of these minimax solutions w.r.t. W_{ijk}.

DEFINITION 4. A d.r. D defined by $\phi(x)$ is said to be an unlikelihood ratio d.r. if there exist non-negative constants a_{ij} ($i \neq j$; $i = 1, \cdots, l$; $j = 1, \cdots, m$), where for each i at least one $a_{ij} > 0$, such that for any k and any x for which $\phi_k(x) > 0$, $\sum_{i \neq k} a_{ik} f_i^n(x) \leqq \sum_{i \neq j} a_{ij} f_i^n(x)$ for all $j \neq k$.
Setting $a_{ij} = \xi_{ij} / \beta_{ij}$, where $\xi = (\xi_{11}, \xi_{12}, \cdots, \xi_{lm})$ denotes an a priori distribution over Ω', we have from Wald's Theorem 5.1 that any Bayes d.r. relative to ξ is an unlikelihood ratio d.r. and conversely. Lindley [10] introduced such d.r.'s, obtained by his "method of minimum unlikelihood." Hereafter, we shall suppose $\xi_{ij} = 0$ for every i, j for which $\beta_{ij} = \infty$ without loss of generality.

Wald's Theorem 5.3 asserts the existence of a least favorable distribution ξ^0, and that any Bayes d.r. relative to ξ^0 is a minimax d.r. and conversely; moreover,

$$p_{ij}(D^0)/\beta_{ij} = \max_{i,j} [p_{ij}(D^0)/\beta_{ij}] \qquad \text{for any } i, j \text{ for which } \xi_{ij}^0 > 0. \tag{11}$$

Apparently, however, there are no general conditions under which all $\xi_{ij}^0 > 0$, and consequently we have no proof of the admissibility of a minimax d.r. In fact, supposing $l = m$ and the β_{ij}'s satisfy $\sum_j^m \beta_{ij} = 1$ for every i, then $\xi_{ij}^0 > 0$ for all i, j would imply $p_{ij} = \beta_{ij}$, regardless of the sample size! Geometrically, the convex body in the $l \cdot m$-dimensional space with coordinate axes p_{ij}, corresponding to all possible d.r.'s for a fixed sample size, is not necessarily intersected by the line determined by $p_{ij}/\beta_{ij} = p_{i'j'} / \beta_{i'j'}$ for all pairs of subscripts corresponding to incorrect decisions. However, we do have the following theorem in this regard, assuming $l = m$ and A_i is "correct" when f_i is true ($i = 1, \cdots, m$).

THEOREM 5. *Suppose Assumption 1 holds and that $\sum_{j \neq i} \beta_{ij} < 1$ for every i. For any sample size greater than or equal to the M.E. sample size, a least favorable distribution ξ^0 has the property $\sum_{i=1}^m \xi_{ij}^0 > 0$ for every j.*
The theorem may be proved by a contradiction, using Assumption 1, Definition 4, Lemma 1, and constructing a Bayes d.r. relative to ξ^0. From Theorem 5 and (11), it follows that $p_{ij}(D_N^0)/\beta_{ij}$ attains its maximum for at least one value of i for every j, where D_N^0 is a minimax d.r. for samples of the M.E. size.

Example 2. We shall consider unlikelihood ratio d.r.'s for samples of size n for Example 1 above. For simplicity, suppose $\sigma = 1$, $l = m = 3$, and $\theta_2 = 0$, the alternatives A_1, A_2, A_3 corresponding respectively to the densities f_1, f_2, f_3.

A d.r. with acceptance regions

$$R_1^n = \{x : h_1^n(x) \leqq h_2^n(x), h_1^n(x), \leqq h_3^n(x)\},$$
$$R_2^n = \{x : h_2^n(x) < h_1^n(x), h_2^n(x) \leqq h_3^n(x)\},$$
$$R_3^n = \{x : h_3^n(x) < h_1^n(x), h_3^n(x) < h_2^n(x)\},$$

where $h_1^n(x) = a_{ji}f_j^n + a_{ki}f_k^n$ and (i, j, k) is a permutation of $(1, 2, 3)$, is an unlikelihood ratio d.r. for the weights (a_{ij}). Denoting the sample mean by $\bar{x}$, we may replace h_i^n by $g_i^n = [a_{ji} \exp (n\theta_j\bar{x} - n\theta_j^2/2) + a_{ki} \exp (n\theta_k\bar{x} - n\theta_k^2/2)]$. Now g_1^n is an increasing function of $\bar{x}$ and g_3^n a decreasing function; g_2^n has a single stationary point, a minimum. By sketching the three g_i^n functions, it is clear that if none of the acceptance regions is to be empty, one of three possibilities must obtain: the acceptance regions are of the form $R_1 = \{x:\bar{x} \leqq c_1 \text{ or } c_3 \leqq \bar{x} \leqq c_4\}$, $R_2 = \{x:c_2 \leqq \bar{x} \leqq c_3\}$, $R_3 = \{x:c_1 \leqq \bar{x} \leqq c_2 \text{ or } \bar{x} \geqq c_4\}$, where either $c_1 = c_2$, or $c_3 = c_4$, or both. (Equality signs have been assigned everywhere in the R_i's for simplicity.) Let c $(=2$ or $3)$ denote the number of c_i's to be determined. The c_i's may be obtained by solving $c + 1$ of the six equations $p_{ij} = \rho\beta_{ij}$ for the c_i's and ρ, the choice of the equations to be solved being such that $p_{ij} \leqq \rho\beta_{ij}$ for all six pairs of subscripts. Theorem 5 may be helpful in this choice of equations. To obtain an M.E. d.r., the sample size n is to be minimized subject to $\rho = \rho_n \leqq 1$. Similar methods may be applied to simple discrimination problems concerning any distribution of the exponential type.

2. Composite discrimination.

A. *The problem.* In this section we allow a continuum of possible density functions. For specificity, assume Ω to be the space of a real- or vector-valued parameter θ indexing the class of density functions w.r.t. μ with elements $f(x, \theta)$.

We further suppose that disjoint subsets $\omega_1, \cdots, \omega_l$ of Ω are specified such that for every pair i, j $(i = 1, \cdots, m; j = 1, \cdots, l)$ there is a definite preference for or against the decision A_j if the true $\theta \varepsilon \omega_i$. We suppose that none of the decisions is definitely preferred if θ is not in some ω_i; this "indifference region" is excluded from Ω for convenience. Under these assumptions, we say that the corresponding decision problem is one of "composite discrimination" and a d.r. is a d.r. for "discriminating among $\omega_1, \cdots, \omega_l$." A d.r. $D = D_n$ is characterized by the functions

$$p_j(\theta, D) = \Pr\ (D \text{ chooses } A_j|\theta) = \int_{\mathfrak{X}} \phi_j(x)f^n(x, \theta)d\mu^n \quad (j = 1, \cdots, m),$$

defined for all $\theta \varepsilon \Omega$.

We again consider two criteria for choosing a d.r. for composite discrimination. The first requires $l = m$ and A_i to be a "correct" decision if $\theta \varepsilon \omega_i$ and "incorrect" if $\theta \varepsilon \omega_j$ $(j \neq i)$. For the second criterion, we suppose that corresponding to each ω_i one or more alternatives A_j is preferable when $\theta \varepsilon \omega_i$.

The definitions and comments of Section 1.A may be restated, substituting only ω_i for f_i, $\inf_{\theta\varepsilon\omega_i} p_i(\theta, D)$ for $p_i(D)$, and $\sup_{\theta\varepsilon\omega_i} q_i(\theta, D)$ for $q_i(D)$. When $l = m = 2$, an M.E. 2-d.r. may be considered as a test of the hypothesis that $\theta \varepsilon \omega_1$ against the class of alternatives $\theta \varepsilon \omega_2$, satisfying bounds on the two kinds of error; such a d.r. may be obtained by considering, for each n, tests of size $1 - \alpha_1$ w.r.t. ω_1, which maximize the minimum power w.r.t. ω_2 and choosing that test for which n is a minimum subject to the minimum power being at least α_2 [7].

Before extending this result to m-d.r.'s for composite discrimination, we require some results in minimax decision theory.

B. *Minimax decision rules for fixed sample sizes.* We prove three theorems which may be useful in finding minimax d.r.'s. Also, if a sufficient statistic with a monotone likelihood ratio exists, Karlin and Rubin's complete class theorem may be applicable [1], [8]. Sverdrup's results [13] should also be noted.

We shall use a number of Wald's results in [14], Section 3.5 and 5.1.4, with some alteration in his assumptions and notation. We denote a weight function by $W(\theta, A_j) = W_j(\theta)$ $(j = 1, \cdots, m)$ and the corresponding risk function when using a d.r. D by $r(\theta, D)$. An a priori distribution over the Borel subsets $\{\omega\}$ of Ω is denoted by $\Xi = (\xi, \lambda)$, where $\Xi(\omega) = \Pr\ (\theta \,\varepsilon\, \omega) = \sum_{i=1}^{l} \xi_i \lambda_i(\omega)$ and $\xi_i = \Xi(\omega_i)$, $\lambda_i(\omega) = \Pr\ (\theta \,\varepsilon\, \omega | \theta \,\varepsilon\, \omega_i)$ $(i = 1, \cdots, l)$. The average risk relative to Ξ is denoted by $r(\Xi, D)$. Other terminology and notation will be self-evident. Wald's Assumptions 5.1 and 5.6, his remarks on page 148 characterizing a Bayes solution, and his theorems 5.11, 5.12, 3.8, 3.9, and 3.10 characterizing minimax solutions are especially pertinent to what follows. Lehmann's existence theorem for least favorable distributions [9] might also be noted.

ASSUMPTION 2. *For each i, j pair $(i = 1, \cdots, l; j = 1, \cdots, m)$, $W_j(\theta)$ equals a constant, say W_{ij}, for all $\theta \,\varepsilon\, \omega_i$.*

(That is, for each alternative, the loss varies only from subset to subset among $\omega_1, \cdots, \omega_l$ and not within any subset.) This assumption is sufficient to imply the validity of Wald's Assumptions 3.1 to 3.6 (see his remarks on page 148). For a given set of conditional distributions $\lambda = (\lambda_1, \cdots, \lambda_l)$, we denote

$$f_i^{\lambda}(x) = \int_{\omega_i} f^n(x, \theta)\, d\lambda_i \qquad (i = 1, \cdots, l); \tag{12}$$

n is fixed and need not be evident in the notation.

THEOREM 6. *If Assumption 2 holds, a necessary and sufficient condition for a d.r. D^* to be a Bayes d.r. relative to $\Xi = (\xi, \lambda)$ for discriminating among $\omega_1, \cdots, \omega_l$ is that D^* be a Bayes d.r. relative to ξ for discriminating among $f_1^{\lambda}, \cdots, f_l^{\lambda}$ w.r.t. the weight function W_{ij}. The average risk in the two cases are equal.*

PROOF. Using Assumption 2 and (12), we have

$$\textstyle\int_{\Omega} W_j(\theta) f^n(x, \theta)\, d\Xi = \sum_{i=1}^{l} \xi_i W_{ij} f_i^{\lambda}(x).$$

The first part of the theorem follows immediately, using Wald's Theorem 5.1 and second paragraph on page 148. By expressing $r(\theta, D)$ as in Wald's (5.81), interchanging the order of integration, using (12) and Wald's (5.2), we have for any d.r. D,

$$\int_{\omega_i} r(\theta, D)\, d\lambda_i = r(f_i^{\lambda}, D). \tag{13}$$

Denoting by $r_{\lambda}(\xi, D)$ the average risk relative to ξ when discriminating among $f_1^{\lambda}, \cdots, f_l^{\lambda}$, we have

$$r(\Xi, D) = r_{\lambda}(\xi, D), \tag{14}$$

completing the proof.

THEOREM 7. *Suppose Assumption 2 holds. Necessary and sufficient conditions that* $\Xi^0 = (\xi^0, \lambda^0)$ *be a least favorable distribution and* D^0 *a minimax d.r. for discriminating among* $\omega_1, \cdots, \omega_l$ *are that*

(i) ξ^0 *is a least favorable distribution and* D^0 *is a minimax d.r. w.r.t.* W_{ij} *for discriminating among* $f_1^{\lambda^0}, \cdots, f_l^{\lambda^0}$; *and*

(ii) *for any* i *for which* $\xi_i^0 > 0$, $\int_{\omega_i} r(\theta, D^0)\, d\lambda_i^0 = \sup_{\theta\epsilon\omega_i} r(\theta, D^0)$. *Moreover, the maximum risks in the two cases are equal; i.e.,*

$$\sup_{\Omega} r(\theta, D^0) = \max_{1 \le i \le l} r(f_i^{\lambda^0}, D^0). \tag{15}$$

PROOF. Necessity: Since Ξ^0 is least favorable, $\inf_D r(\Xi^0, D) \ge \inf_D r((\xi, \lambda^0), D)$ for any ξ, so that, using (14), $\inf_D r_{\lambda^0}(\xi^0, D) \ge \inf_D r_{\lambda^0}(\xi, D)$; that is, ξ^0 is least favorable. Using Wald's Theorem 3.9 and then Theorem 6, D^0 is a Bayes d.r. relative to Ξ^0 and a minimax d.r. for discriminating among $f_1^{\lambda^0}, \cdots, f_l^{\lambda^0}$.

We shall now verify (15). Using Wald's Theorem 5.3 (iii), $\max_i r(f_i^{\lambda^0}, D^0) = \sum_i \xi_i^0 r(f_i^{\lambda^0}, D^0) = r_{\lambda^0}(\xi^0, D^0)$, so that together with (14) and Wald's Theorem 3.10, we have $\max_i r(f_i^{\lambda^0}, D^0) = r(\Xi^0, D^0) = \sup_\Omega r(\theta, D^0)$. Continuing with the necessity, for any i for which $\xi_i^0 > 0$, we have $r(f_i^{\lambda^0}, D^0) = \max_i r(f_i^{\lambda^0}, D^0)$ and $\sup_{\omega_i} r(\theta, D^0) = \sup_\Omega r(\theta, D^0)$ by Wald's Theorem 3.10, which, together with (15) and (13), prove (ii).

Sufficiency: By Wald's Theorem 3.9 and Theorem 6, D^0 is a Bayes d.r. relative to $\Xi^0 = (\xi^0, \lambda^0)$; i.e., $r(\Xi^0, D^0) = \inf_D r(\Xi^0, D)$. Hence, we need only prove that Ξ^0 is a least favorable distribution. Suppose it is not; then there exists a $\Xi = (\xi, \lambda)$ such that $\inf_D r(\Xi^0, D) < \inf_D r(\Xi, D)$. But $\inf_D r(\Xi, D) \le r(\Xi, D^0) = \sum_i \xi_i \int_{\omega_i} r(\theta, D^0)\, d\lambda_i \le \sum_i \xi_i \sup_{\omega_i} r(\theta, D^0) \le \sup_\Omega r(\theta, D^0)$. Combining these last three results, $r(\Xi^0, D^0) < \sup_\Omega r(\theta, D^0)$.

By Wald's Theorem 5.3 (iii), for any i for which $\xi_i^0 > 0$, $r(f_i^{\lambda^0}, D^0) = \max_i r(f_i^{\lambda^0}, D^0)$, which, together with (13) and (ii), implies $\sup_{\omega_i} r(\theta, D^0) = \max_i \sup_{\omega_i} r(\theta, D^0) = \sup_\Omega r(\theta, D^0)$. Hence, from (ii),

$$r(\Xi^0, D^0) = \sum_i \xi_i^0 \int_{\omega_i} r(\theta, D^0)\, d\lambda_i^0 = \sup_\Omega r(\theta, D^0),$$

a contradiction. Q.E.D.

THEOREM 8. *Suppose Assumption 2 holds, and suppose* $\{\lambda^\nu\}$ *is a sequence of sets of conditional a priori distributions and* D^0 *a d.r. such that*

$$\lim_{\nu\to\infty} \int_{\omega_i} r(\theta, D^\nu)\, d\lambda_i^\nu = \sup_{\omega_i} r(\theta, D^0) \qquad (i = 1, \cdots, l), \tag{16}$$

where for each $\nu = 1, 2, \cdots$, D^ν *is a minimax d.r. for discriminating among* $f_1^{\lambda^\nu}, \cdots, f_l^{\lambda^\nu}$. *Then* D^0 *is a minimax d.r. for discriminating among* $\omega_1, \cdots, \omega_l$.

PROOF. By Wald's Theorem 5.3, for each ν there exists a least favorable distribution ξ^ν, and D^ν is a Bayes d.r. relative to ξ^ν for discriminating among $f_1^{\lambda^\nu}, \cdots, f_l^{\lambda^\nu}$; i.e., for any d.r. D, $r_{\lambda^\nu}(\xi^\nu, D^\nu) \le r_{\lambda^\nu}(\xi^\nu, D)$, and hence, using (14),

$$\sum_i \xi_i^\nu \int_{\omega_i} r(\theta, D^\nu)\, d\lambda_i^\nu \le \sum_i \xi_i^\nu \int_{\omega_i} r(\theta, D)\, d\lambda_i^\nu \le \sup_\Omega r(\theta, D) \tag{17}$$

Now each sequence $\{\xi_i^r\}$ has at least one limit point; let $\{\Xi^{r_j}\}$, $j = 1, 2, \cdots$, be a sub-sequence of $\{\Xi^r = (\xi^r, \lambda^r)\}$ for which each $\xi_i^{r_j}$ converges to a limit, say ξ_i^0; then $\sum_i \xi_i^0 = 1$. By Wald's Theorem 5.3 (iii) and (13), for each i for which $\xi_i^r > 0$, $\int_{\omega_i} r(\theta, D^r)\, d\lambda_i^r = \max_i \int_{\omega_i} r(\theta, D^r)\, d\lambda_i^r$ so that, from (16), for each i for which $\xi_i^0 > 0$, $\sup_{\omega_i} r(\theta, D^0) = \max_i \sup_{\omega_i} r(\theta, D^0) = \sup_\Omega r(\theta, D^0)$. Hence, from (16), $\lim_{j=\infty} \sum_i \xi_i^{r_j} \int_{\omega_i} r(\theta, D^{r_j})\, d\lambda_i^{r_j} = \sum_i \xi_i^0 \sup_{\omega_i} r(\theta, D^0) = \sup_\Omega r(\theta, D^0)$, which, together with (22), asserts $\sup_\Omega r(\theta, D^0) \leqq \sup_\Omega r(\theta, D)$ for any D. Q.E.D.

If a least favorable distribution exists, the problem reduces to one of simple discrimination, so that if μ is non-atomic only non-randomized d.r.'s need be considered. A lemma for the case of composite discrimination analogous to Lemma 1 may be derived.

C. *Most economical decision rules relative to a vector* α. As in Section 1.B, we shall apply the above theory to two specific weight functions $W_j(\theta)$ and develop in each case a method of obtaining M.E. d.r.'s relative to α. We assume $l = m$. First, let

$$W(\theta, A_j) = W_j(\theta) = -1/\alpha_j \qquad \text{if } \theta \in \omega_j \text{ and } 0 \text{ otherwise.}[6] \tag{18}$$

The risk w.r.t. $W_j(\theta)$ is $r(\theta, D) = -p_i(\theta, D)/\alpha_i$ if $\theta \in \omega_i$ ($i = 1, \cdots, m$), and $\sup_{\omega_i} r(\theta, D) = -\inf_{\omega_i} p_i(\theta, D)/\alpha_i$ ($i = 1, \cdots, m$). By Wald's Theorem 5.12 (i), there exists a minimax d.r. D^0 for any (fixed) sample size.

THEOREM 9. *For each* $n = 0, 1, 2, \cdots$, *let* D_n^0 *be a minimax d.r. w.r.t. the weight function* (18) *for samples of fixed size n. Suppose for some n,* $\sup_\Omega r(\theta, D_n^0) \leqq -1$ *and let N be the least such integer. Then* D_n^0 *is an M.E. d.r. relative to* α *for discriminating among* $\omega_1, \cdots, \omega_m$. *Conversely, if there exists an M.E. d.r. relative to* α *for discriminating among* $\omega_1, \cdots, \omega_m$, *and the M.E. sample size is N, then* D_N^0 *is an M.E. d.r.*

The proof is like that of Theorem 1, replacing $p_i(D_n)$ by $\inf_{\omega_i} p_i(\theta, D_n)$.

Note that (18) satisfies Assumption 2 with W_{ij} given by (3). Hence, if a least favorable distribution $\Xi^0 = (\xi^0, \lambda^0)$ exists, Theorems 6 and 7 imply that the composite discrimination problem may be treated as a simple discrimination problem with $f_i(x) = f_i^{\lambda^0}(x) = \int_{\omega_i} f(x, \theta)\, d\lambda_i^0$, and the theory of Section 1 will be applicable. If a least favorable distribution does not exist, Theorem 8 asserts that by a similar treatment for a sequence of a priori distributions having certain properties in the limit, it may be possible to solve the composite discrimination problem. Now suppose a least favorable distribution $\Xi^0 = (\xi^0, \lambda^0)$ exists. By Theorem 7(ii),

$$\int_{\omega_i} p_i(\theta, D^0)\, d\lambda_i^0 = \inf_{\theta \in \omega_i} p_i(\theta, D^0) \qquad \text{for any } i \text{ for which } \xi_i^0 > 0.$$

ASSUMPTION 3. If R is a subset of $\mathfrak{X}$ for which $\int_R f^n(x, \theta)\, d\mu^n = 0$ for some $\theta \in \Omega$, then $\int_R f^n(x, \theta)\, d\mu^n = 0$ for all $\theta \in \Omega$.

This assumption implies Assumption 1 for the density functions $f_1^\lambda, \cdots, f_m^\lambda$,

[6] See footnote 3.

defined by (12), for any set of conditional distributions λ. If Assumption 3 holds, and if a least favorable distribution exists, it follows from Theorem 2, Wald's Theorem 5.3(iii) and (18) that

$$\frac{1}{\alpha_1} \inf_{\theta \varepsilon \omega_1} p_1(\theta, D^0) = \cdots = \frac{1}{\alpha_m} \inf_{\theta \varepsilon \omega_m} p_m(\theta, D^0), \tag{19}$$

where D^0 is a minimax d.r.

As a second approach, consider the weight function:

$$W(\theta, A_j) = W_j(\theta) = 1/\beta_i \qquad \text{if } \theta \varepsilon \omega_i,\ i \neq j, \text{ and } 0 \text{ otherwise}, \tag{20}$$

where $\beta_i = 1 - \alpha_i$ as before. Then $r(\theta, D) = q_i(\theta, D)/\beta_i$ if $\theta \varepsilon \omega_i$ $(i = 1, \cdots, m)$. We may proceed analogously to the first approach, making changes corresponding to those made analogously in Section 1. We thus obtain a theorem analogous to Theorem 9 and also

THEOREM 10. *Suppose Assumption 3 holds and that a least favorable distribution exists. For any sample size greater than or equal to the M.E. sample size,*

$$\frac{1}{\beta_1} \sup_{\theta \varepsilon \omega_1} q_1(\theta, D^0) = \cdots = \frac{1}{\beta_m} \sup_{\theta \varepsilon \omega_m} q_m(\theta, D^0) \tag{21}$$

where D^0 is a minimax d.r.

No proof of admissibility of the M.E. d.r.'s derived in this section has been obtained. However, if Assumption 3 holds and there exists a least favorable distribution, it can easily be verified that there does not exist a d.r. D_N' for which $\inf_{\omega_i} p_i(\theta, D_N') \geq \inf_{\omega_i} p_i(\theta, D_N^0)$ $(i = 1, \cdots, m)$ with strict inequality for at least one i, where D_N^0 is an M.E. d.r. obtained by either of the minimax methods.

D. *Most economical decision rules relative to a matrix β.* Just as the approach of Section 1.B was extended in Section 1.C, we shall extend the approach of Section 2.C in this section to the consideration of M.E. d.r.'s for composite discrimination relative to $\beta = (\beta_{ij})$.

Suppose n is fixed, and consider parameter spaces $\Omega_1, \cdots, \Omega_m$, each Ω_j being identical to Ω, and denote $\Omega' = U_j\Omega_j$. For each j, denote the corresponding subsets by $\omega_{1j}, \cdots, \omega_{lj}$. Define a weight function $W(\theta, A_k) = W_k(\theta)$ for $k = 1, \cdots, m$, by

$$W_k(\theta) = 1/\beta_{ij} \quad \text{if } \theta \varepsilon \omega_{ij} \text{ and } j = k\ (i = 1, \cdots, l; j = 1, \cdots, m), \text{ and } 0 \text{ otherwise}. \tag{22}$$

Then $r(\theta, D) = p_j(\theta, D)/\beta_{ij}$ if $\theta \varepsilon \omega_{ij}$. Let Ξ be an a priori distribution over Ω' with components $\xi_{ij} = \Xi(\omega_{ij})$ and $\lambda_{ij}(\omega) = \Pr(\theta \varepsilon \omega \mid \theta \varepsilon \omega_{ij})$. For a given set of λ's, denote

$$g_{ij}^{\lambda}(x) = \int_{\omega_{ij}} f^n(x, \theta)\, d\lambda_{ij}. \tag{23}$$

Theorem 9 may be restated and proved, substituting only (22) for (18), $+1$ for -1, and β for α. The theorems of Section 2.B may be applied to obtain mini-

max d.r.'s for composite discrimination w.r.t. the weight function (22) by replacing l in the theorems by $l' = l \cdot m$ and replacing single subscripts i by ij and f_i^λ by g_{ij}^λ. If a least favorable distribution exists, then the composite discrimination problem reduces to a problem of simple discrimination among the "average" density functions g_{ij}^λ defined by (23) w.r.t. a set of "least favorable conditional distributions" λ, and Theorem 5 and the remarks of Section 1.C are applicable. Thus, this method of solution gives unlikelihood ratio d.r.'s as M.E. d.r.'s. If a least favorable distribution does not exist, then a minimax d.r. will be a Bayes d.r. in the wide sense and Theorem 8 may be applicable.

Example 3. Suppose $f(x, \theta)$ is a normal density function with variance σ^2 (known) and mean θ, and

$$\omega_1 = \{\theta : \theta \leqq \theta_1\}, \qquad \omega_2 = \{\theta : \theta_2' \leqq \theta \leqq \theta_2''\}, \qquad \omega_3 = \{\theta : \theta \geqq \theta_3\},$$

for some specified $\theta_1 < \theta_2' \leqq \theta_2'' < \theta_3$. It may be shown that the least favorable conditional distributions over ω_1, ω_2, ω_3 (Theorem 7) assign probability one to θ_1, θ_2, θ_3, where $\theta_2 = \theta_2'$ or θ_2'' determined below. Thus, this example reduces to Example 1. (Karlin and Rubin's results [8] also imply that a minimax rule will be monotone in $\bar{x}$; determining the explicit form of the monotone rule is equivalent to showing that the above distribution is least favorable.)

θ_2 is determined as follows:

$$\begin{aligned} \theta_2 &= \theta_2' \quad \text{if} \quad p_2(\theta_2', D') \leqq p_2(\theta_2'', D'), \\ \theta_2 &= \theta_2'' \quad \text{if} \quad p_2(\theta_2'', D'') < p_2(\theta_2', D''), \end{aligned} \tag{24}$$

where D' and D'' are the solutions to the corresponding simple discrimination problems with $\theta_2 = \theta_2'$ or θ_2'' for fixed n. We shall show that such a determination of θ_2 is complete and consistent by showing that if $p_2(\theta_2'', D'') > p_2(\theta_2', D'')$ then $p_2(\theta_2'', D') > p_2(\theta_2', D')$ and conversely. From (9), with either a prime or double-prime on ρ, D, c_1, and c_2, we have $c_1 = \theta_1 + \sigma\Phi^{-1}(\alpha_1\rho)/\sqrt{n}$ and

$$c_2 = \theta_3 + \sigma\Phi^{-1}(1 - \alpha_3\rho)/\sqrt{n},$$

where $\Phi^{-1}(x) = t$ is defined by $\Phi(t) = x$. Substituting in $p_2(\theta, D)$ it becomes clear that it is a decreasing function of ρ for fixed θ. Now $\alpha_2\rho' = p_2(\theta_2', D')$ and $\alpha_2\rho'' = p_2(\theta_2'', D'')$ so that

$$\alpha_2(\rho'' - \rho') = p_2(\theta_2'', D'') - p_2(\theta_2', D'). \tag{25}$$

Suppose $p_2(\theta_2'', D'') > p_2(\theta_2', D'')$; substituting in (25), it follows that $\rho'' > \rho'$ since p_2 is a decreasing function of ρ. For the same reasons,

$$0 < \alpha_2(\rho'' - \rho') < p_2(\theta_2'', D') - p_2(\theta_2', D').$$

Conversely, in the same manner, if $p_2(\theta_2'', D') > p_2(\theta_2', D')$, then

$$\alpha_2(\rho'' - \rho') > p_2(\theta_2'', D'') - p_2(\theta_2'', D'),$$

and ρ'' must be greater than ρ'; hence,

$$0 < \alpha_2(\rho'' - \rho') < p_2(\theta_2'', D'') - p_2(\theta_2', D'').$$

Other examples with exponential density functions may be treated analogously, and also similar examples for Section 2.D.

Example 4. Now suppose σ is also unknown; denote the mean by μ and replace θ in the ω_i's defined in Example 3 by μ/σ.

Denoting Student's ratio by t and the sample sum of squares by s^2, (t, s) is sufficient for $\theta = (\mu, \sigma)$. If we invoke invariance (under changes in scale), it follows from Blackwell and Girshick's work [1] that a minimax invariant rule must be monotone in t. Theorem 8.8.1 in [1] proves, for the m-decision case as well as the 2-decision case, that invariance is no restriction when discriminating among $\theta_1, \cdots, \theta_m$, where $\theta = \mu/\sigma$. Thus a minimax d.r. for discriminating among $\theta_1, \theta_2, \theta_3$ is monotone in t. By showing that the risk for a monotone rule is a maximum in ω_i at $\mu/\sigma = \theta_i$ (with θ_2 determined as in Example 3), it will follow that a monotone rule in t, with c_1, c_2 and ρ determined by equations of the form (9) with the Φ's replaced by non-central t distribution functions, will be minimax for discriminating among $\omega_1, \omega_2, \omega_3$.

Alternatively, this same result may be obtained by an application of our Theorem 8, letting λ_i^ν assign probability one to sets of (μ, σ) in which

$$\mu/\sigma = \theta_i$$

and letting σ^{-2} be distributed as χ^2 with degrees of freedom tending to 0 as $\nu \to \infty$. The details appear in [3], adapted from a 2-d.r. argument by Hoeffding.

Example 5. We shall derive a three-decision extension of the sign test for the median of an arbitrary distribution function by adapting an example of Hoeffding [6]. (See also [12].) Analogously, an M.E. d.r. concerning any quartile of an arbitrary distribution may be derived.

Let Ω be the class of all density functions f w.r.t. a fixed measure μ on the real line such that $\mu\{x \leqq 0\} > 0$, $\mu\{x > 0\} > 0$. Denote $\theta(f) = \int_{-\infty}^{0} f(x)\, d\mu$. Given $\theta_1, \theta_2', \theta_2'', \theta_3$ $(0 < \theta_1 < \theta_2' \leqq \frac{1}{2} \leqq \theta_2'' < \theta_3 < 1)$, let $\omega_1 = \{f:\theta(f) \leqq \theta_1\}$, $\omega_2 = \{f:\theta_2' \leqq \theta(f) \leqq \theta_2''\}$, $\omega_3 = \{f:\theta(f) \geqq \theta_3\}$. The alternatives A_1, A_2, A_3, corresponding to $\omega_1, \omega_2, \omega_3$, might be that the median of the unknown distribution is "appreciably" less than zero, "close" to zero, "appreciably" greater than zero, respectively.

Let $f(x, \theta) = \theta^{b(x)}(1 - \theta)^{1-b(x)}/c$ if $x \leqq c$ and 0 otherwise where c is an arbitrary positive constant and $b(x) = 1$ if $x \leqq 0$ and 0 otherwise, and let $\lambda = (\lambda_1, \lambda_2, \lambda_3)$ be a set of conditional distributions over $\omega_1, \omega_2, \omega_3$, respectively, where λ_i assigns probability 1 to $f(x, \theta_i)$ and where θ_2 is to be determined as in Example 3. It is easily verified that a minimax d.r. D_n (n fixed) for discriminating among $f_1^\lambda, f_2^\lambda, f_3^\lambda$ is monotone in $t(x) = \sum_k b(x_k)$, the number of non-positive observations, with c_1, c_2 and values of ϕ_i when $t = c_1$ or c_2 determined so that $p_i(\theta_i, D_n) = a_i\rho$ $(i = 1, 2, 3)$ for some ρ; and

$$p_1(\theta, D_n) = B(c_1 - 1) + a_1 b(c_1), \qquad p_3(\theta, D_n) = 1 - B(c_2) + (1 - a_2)b(c_2),$$

$$p_2(\theta, D_n) = B(c_2 - 1) + a_2 b(c_2) - B(c_1) + (1 - a_1)b(c_1),$$

where $B = B_{n,\theta}$ and $b = b_{n,\theta}$ denote the binomial distribution function and probability function, respectively, and $a_i = \phi_i(c_i) = 1 - \phi_{i+1}(c_i)$. (It may be shown that D_n defined above is also minimax for discriminating among

$$b_{n,\theta_1}, \qquad b_{n,\theta_2}, \qquad b_{n,\theta_3}.)$$

This λ may be shown to be least favorable, and an M.E. d.r. may be obtained according to Theorem 9 (see Example 1).

3. Acknowledgments. The author wishes to express his sincere thanks to Professor Wassily Hoeffding, whose stimulating lectures inspired this research, and whose criticisms and suggestions have been invaluable in accomplishing it.

REFERENCES

[1] D. Blackwell and M. A. Girshick, *Theory of Games and Statistical Decisions*, John Wiley and Sons, New York, 1954.

[2] A. Dvoretzky, A. Wald and J. Wolfowitz, "Elimination of randomization in certain statistical decision procedures and zero-sum two-person games," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 1-21.

[3] W. J. Hall, "Most economical multiple-decision rules," Institute of Statistics Mimeograph Series No. 115, University of North Carolina, 1954.

[4] W. J. Hall, "An optimum property of Bechofer's single-sample multiple-decision procedure for ranking means and some extensions," Institute of Statistics Mimeograph Series No. 118, University of North Carolina, 1954.

[5] W. J. Hall, "On uniformly consistent sequences of multiple-decision rules," abstract, *Ann. Math. Stat.*, Vol. 27 (1956), p. 221.

[6] Wassily Hoeffding, " 'Optimum' nonparametric tests," *Second Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, 1951, pp. 83-92.

[7] Wassily Hoeffding, "Lectures on estimation and testing of hypotheses," University of North Carolina, fall quarter 1952 (unpublished).

[8] S. Karlin and H. Rubin, "The theory of decision procedures for distributions with monotone likelihood ratio," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 272-299.

[9] E. L. Lehmann, "On the existence of least favorable distributions," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 408-416.

[10] D. V. Lindley, "Statistical inference," *J. Royal Stat. Soc.*, Series B, Vol. 15 (1953), pp. 30-76.

[11] C. R. Rao, *Advanced Statistical Methods in Biometric Research*, John Wiley and Sons, New York, 1952, p. 311.

[12] Erik Ruist, "Comparison of tests for non-parametric hypotheses," *Arkiv for Matematik*, Vol. 3 (1954), pp. 133-163.

[13] Erling Sverdrup, "Weight functions and minimax procedures in the theory of statistical inference," *Archiv for Mathematik og Naturvidenskab*, Vol. 51 (1952), pp. 1-73.

[14] Abraham Wald, *Statistical Decision Functions*, John Wiley and Sons, New York, 1950.

[15] A. Wald and J. Wolfowitz, "Characterization of the minimal complete class of decision functions when the number of distributions and decisions is finite," *Second Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, 1951, pp. 149-157.

ESTIMATION OF THE MEANS OF DEPENDENT VARIABLES

By Olive Jean Dunn[1]

Statistical Laboratory, Iowa State College

1. Summary. Methods are given for constructing sets of simultaneous confidence intervals for the means of variables which follow a multivariate normal distribution.

In section (3), a set of confidence intervals is obtained for each of two special cases; first when the variances are assumed to be known, and second when the variances are assumed to be equal. These two sets have the property that the confidence is known exactly, rather than merely being bounded below. In the case of known variances, the intervals are of fixed lengths (i.e., the lengths are the same from sample to sample); when the variances are unknown, the intervals are of variable lengths. It may be surprising to note that nothing need be known about the covariances in order to obtain confidence intervals of fixed lengths whose confidence coefficient is exact. These intervals are long, and do not make use of all the information provided by the sample.

Each of sections (4) to (7) considers a different method for obtaining confidence intervals of bounded confidence level. In each section a set of fixed lengths is obtained when the variances are assumed to be known, while a set of variable lengths is obtained when the variances are unknown but equal. In section (5) the set of variable lengths applies to the general multivariate normal distribution, all the other confidence intervals in this paper require some assumption concerning the variances.

In section (8) the sets of intervals are compared on the basis of length. One of the bounded confidence level methods, which has been established only for two or three variables or for an arbitrary number of variables with a special type of correlation matrix, is shown to yield the best possible set. Another of the bounded confidence level methods, whose use is established in general, is shown to be almost as good as the best set for confidence coefficients of practical interest.

It is interesting to notice that intervals with bounded confidence level, are found which are much shorter than the ones whose confidence level is exact. This need not surprise us, however. In the case of just one variable, we might easily find that the 95% confidence intervals for the mean using the t-statistic were shorter on the average than 94% confidence intervals using order statistics. Moreover, since in admitting sets of confidence intervals with bounded confidence level we consider a much broader class of methods, we might almost expect that some of them would give better intervals.

2. Introduction. The problem of estimating the unknown means of dependent variables arises frequently in situations where repeated measurements are made

Received February 15, 1957; revised March 25, 1958.

[1] Now at the University of California at Los Angeles.

on the same individuals, and the assumption of independence is unjustifiable. In biological research, for example, growth data are often obtained with measurements taken on n individuals at k different times; the measurements would be highly correlated. The psychologist might measure n individuals' responses to k different levels of a stimulus; again, a high degree of dependence would be expected. The point estimates chosen for the means would be the same as for independent variables; in this paper we wish to develop simultaneous confidence intervals for the means.

Let $y_1, \cdots, y_k$ be k jointly distributed variables whose means are $\mu_1, \cdots, \mu_k$ respectively. A set of simultaneous confidence intervals for $\mu_1, \cdots, \mu_k$ with confidence coefficient $1 - \alpha$ consists of $2k$ functions of the sample values, say g_i and h_i, $i = 1, 2, \cdots, k$, with the following property: if E_i is the event that the interval g_i to h_i covers μ_i, $i = 1, 2, \cdots, k$, then the probability that E_1, $E_2, \cdots, E_k$ occur simultaneously is greater than or equal to $1 - \alpha$, where $0 < \alpha < 1$. Symbolically,

$$P(E_1E_2, \cdots, E_k) = P(g_1 < \mu_1 < h_1, \cdots, g_k < \mu_k < h_k) \geqq 1 - \alpha.$$

If the inequality sign holds, the set is of bounded confidence level.

Paul G. Hoel has in a recent paper [1] given a method for estimating a mean regression curve and a confidence band for it which is applicable to the situations we have in mind provided one assumes the existence of a polynomial regression curve of a given degree. In this paper we shall assume that the experimenter is actually interested in the regression curve, but is either unwilling to make the necessary laborious calculations or else is unable to make the necessary assumptions concerning its form. He knows that there exist methods for studying linear contrasts among the means, but this is not what he wishes to do. He might indeed decide to make k different 95% confidence intervals, one for each of the k means; this is satisfactory only when he focuses on one individual mean.

We shall assume, then, that he will welcome a set of k confidence intervals, one for each mean, being assured, with a high probability, that such a set covers all k means simultaneously.

Another situation in which such a set of intervals would be useful arises when a regression line, curve, or surface has been fitted, and several predictions are made on the basis of it.

Suppose, for example, that the assumption has been made that the variables x_i are normally distributed with means $\alpha + \beta t_i$ and variances σ^2, and that the maximum likelihood estimate $\hat{\alpha} + \hat{\beta} t_i$ has been calculated from a sample of size m.

At any particular value of t, say t_0, one can obtain a prediction interval for x_0, an observation drawn at random from the x's belonging to t_0, by using the fact that $u_0 = x_0 - \hat{\alpha} - \hat{\beta} t_0$ is normally distributed. But the research worker is cautioned not to do this for more than one value of t, and of course this is exactly what he wishes to do.

If he goes ahead and gets such intervals at k different points, say $t_1^*, \cdots, t_k^*$, he has the same unsatisfactory situation as with repeated tests of significance. The variables $u_i^* = x_i^* - \hat{\alpha} - \hat{\beta} t_i$, where x_i^* is an observation chosen at random

from the x's at $t = t_i^*$, $i = 1, 2, \cdots, k$, are normally distributed and are correlated; thus the methods of this paper may be used to give simultaneous prediction intervals for the points $x_i^*, \cdots, x_k^*$.

3. Confidence regions using independent linear combinations.

3.1. Assuming first known variances, we seek independent linear combinations of the sample values which can be used to give a set of confidence intervals of fixed lengths whose confidence level is exact.

The observations $y_{1j}, y_{2j}, \cdots, y_{kj}$, $j = 1, \cdots, n$, are a random sample of n observations from $n_k(y_1, \cdots, y_k)$, the multivariate normal distribution with unknown means, $\mu_1, \cdots, \mu_k$, known variances, $\sigma_1^2, \cdots, \sigma_k^2$, and unknown covariances λ_{is}, $i \neq s$.

Let $z_i = \sum_{j=1}^n a_{ji} y_{ij}$, $i = 1, \cdots, k$, with the following restrictions on the a_{ji}:

$$(1) \qquad \sum_{j=1}^n a_{ji} = 1, \qquad i = 1, \cdots, k$$

$$(2) \qquad \sum_{j=1}^n a_{ji} a_{js} = 0, \qquad i \neq s$$

$$(3) \qquad \sum_{j=1}^n a_{ji}^2 = c^2, \qquad i = 1, 2, \cdots, k.$$

The means, variances, and covariances of the z_i may then be calculated, remembering that $E(y_{ij} - \mu_i)(y_{sj} - \mu_s) = \lambda_{is}$, but that (since two observations in a random sample are independent) $E(y_{ij} - \mu_i)(y_{sj'} - \mu_s) = 0$ for $j \neq j'$. The means of the z_i are calculated to be μ_i, $i = 1, \cdots, k$, their variances are proportional to $\sigma_1^2, \cdots, \sigma_k^2$, and their covariances are zero.

To determine the a_{ji}, let $A = (a_{ji})$, an $n \times k$ matrix. The columns of A may be considered to be k vectors in an n-dimensional Euclidean space, each with an end fixed at the origin. The three conditions imply (1) that the k vectors have their endpoints on the plane which passes through the unit points on the coordinate axes, $P: \sum_{i=1}^n a_i = 1$; (2) that they be mutually orthogonal; and (3) that their lengths equal c.

If $n \geqq k$, the columns of

$$D = \left[\begin{array}{cccc} c & 0 & \cdots & 0 \\ 0 & c & \cdots & 0 \\ \cdots & & & \\ 0 & 0 & \cdots & c \\ \hline 0 & 0 & \cdots & 0 \\ \cdots & & & \\ 0 & 0 & \cdots & 0 \end{array}\right], \text{ an } n \times k$$

matrix, are k mutually orthogonal vectors of length c whose endpoints lie on any plane

$$P': \frac{a_1}{c} + \cdots + \frac{a_k}{c} + \frac{a_{k+1}}{m_{k+1}} + \cdots + \frac{a_n}{m_n} = 1.$$

The plane P' can be rotated into the plane P provided the distances of the two planes from the origin are equal; this will be true if

$$c^2 = \frac{k}{n - \frac{1}{m_{k+1}^2} - \cdots - \frac{1}{m_n^2}}.$$

To make the lengths of the confidence intervals formed from the z_i as small as possible, c^2 should be minimized. This is accomplished by choosing for P' the plane $\sum_{i=1}^{k} (a_i/c) = 1$; then $c^2 = (k/n)$.

The solution is then $A = BCD$, where B is an $n \times n$ orthogonal matrix whose first column consists of the elements $n^{-\frac{1}{2}}$;

$$C = \left[\begin{array}{ccc|ccc} \frac{1}{\sqrt{k}} & \cdots & \frac{1}{\sqrt{k}} & 0 & \cdots & 0 \\ \cdots & & \cdots & & & \\ \cdots & & \cdots & 0 & \cdots & 0 \\ \hline 0 & \cdots & 0 & 0 & \cdots & 0 \\ \cdots & & & & & \\ 0 & \cdots & 0 & 0 & \cdots & 0 \end{array}\right], \text{ an } n \times n$$

matrix consisting of zeros except for a $k \times k$ orthogonal matrix in the upper left corner whose first row is $k^{-\frac{1}{2}}, \cdots, k^{-\frac{1}{2}}$; and D is defined as before, with $c = (k/n)$. For C rotates the column vectors of D into vectors whose endpoints lie on the plane $a_1 = n^{-\frac{1}{2}}$. B^{-1} rotates the plane $\sum_{i=1}^{n} a_i = 1$ into the plane $a_1 = n^{-\frac{1}{2}}$, so that B rotates the k mutually orthogonal vectors of length $(k/n)^{\frac{1}{2}}$ into vectors whose endpoints lie on the plane $\sum_{i=1}^{n} a_i = 1$. The problem thus reduces to that of writing down a $k \times k$ orthogonal matrix and an $n \times n$ orthogonal matrix.

The $z_1, \cdots, z_k$ are then independently normally distributed with means μ_i and variances $(k/n)\sigma_i^2$. Thus

$$P\left(z_1 - \sqrt{\frac{k}{n}}\,\sigma_1 c_\alpha < \mu_1 < z_1 + \sqrt{\frac{k}{n}}\,\sigma_k c_\alpha, \cdots, z_k - \sqrt{\frac{k}{n}}\,\sigma_k c_\alpha < \mu_k < z_k + \sqrt{\frac{k}{n}}\,\sigma_k c_\alpha\right) =$$

$1 - \alpha$, where c_α is defined by

$$N(c_\alpha) = \frac{1 + (1 - \alpha)^{1/k}}{2},$$

with N the cumulative distribution function of the standard normal variable. The set of confidence intervals is $z_i \pm (k/n)^{\frac{1}{2}}\sigma_i c_\alpha$.

3.2. When the variances are unknown but are assumed to be equal, the same method may be used to construct t-variables whose numerators are independent but which have the same denominator, provided $n > k$. Let $\sigma_i^2 = \sigma^2, i = 1, \cdots, k$.

Let

$$z_i = \sum_{j=1}^{n} a_{ji} y_{ij}, \qquad i = 1, \cdots, k,$$

and

$$u_m = \sum_{j=1}^{n} b_{jm} y_{rj}, \qquad m = 1, \cdots, n-k,$$

where r is any integer from 1 to k. The choice of r is arbitrary. It may be the same for each u_m, or different r's may be used for the different values of m. The problem is to determine the a_{ji} and b_{jm} so that $z_1, \cdots, z_k, u_1, \cdots, u_{n-k}$ will be independently normally distributed variables with $E(z_i) = \mu_i$, $i = 1, \cdots, k$; $E(u_m) = 0$, $m = 1, \cdots, n-k$, $E(z_i - \mu_i)^2 = (k/n)\sigma^2$, $i = 1, \cdots, k$; $E(u_m^2) = \sigma^2$, $m = 1, \cdots, n-k$. This will be accomplished provided

(1) $\sum_{j=1}^{n} a_{ji} = 1$, $i = 1, \cdots, k$, since $E(z_i) = \mu_i \sum_{j=1}^{n} a_{ji} = \mu_i$.

(2) $\sum_{j=1}^{n} a_{ji}^2 = \dfrac{k}{n}$, $i = 1, \cdots, k$, since $E(z_i - \mu_i)^2 = \sigma^2 \sum_{j=1}^{n} a_{ji}^2 = \dfrac{k}{n}\sigma^2$.

(3) $\sum_{j=1}^{n} a_{ji} a_{js} = 0$, $i \neq s$, since $E(z_i - \mu_i)(z_s - \mu_s) = \lambda_{is} \sum_{j=1}^{n} a_{ji} a_{js} = 0$.

(4) $\sum_{j=1}^{n} b_{jm} = 0$, $m = 1, \cdots, n-k$, since $E(u_m) = \mu_m \sum_{j=1}^{n} b_{jm} = 0$.

(5) $\sum_{j=1}^{n} b_{ji}^2 = 1$, $i = 1, \cdots, n-k$, since $E(u_m^2) = \sigma^2 \sum_{j=1}^{n} b_{jm}^2 = \sigma^2$.

(6) $\sum_{j=1}^{n} b_{jm} b_{js} = 0$, $m \neq s$, since $E(u_m u_s) = E(y_r - \mu_r)(y_{r'} - \mu_{r'}) \sum_{j=1}^{n} b_{jm} b_{js} = 0$.

(7) $\sum_{j=1}^{n} a_{ji} b_{jm} = 0$, $i = 1, \cdots, k$; $m = 1, \cdots, n-k$, since $E(z_i - \mu_i)(u_m)$

$$= \lambda_{ir} \sum_{j=1}^{n} a_{ji} b_{jm} = 0.$$

Thus n mutually orthogonal vectors are needed, k of length $(k/n)^{\frac{1}{2}}$ with endpoints on the plane $\sum_{i=1}^{n} a_i = 1$, and $n-k$ of length one with endpoints on the plane $\sum_{i=1}^{n} a_i = 0$.

Let

$$D = \begin{array}{r} k\left\{\vphantom{\begin{matrix}\sqrt{\frac{k}{n}}\\ \sqrt{\frac{k}{n}}\\ \cdots \\ \sqrt{\frac{k}{n}}\end{matrix}}\right. \\ n-k\left\{\vphantom{\begin{matrix}0\\0\\ \cdots\\0\end{matrix}}\right. \end{array} \left[\begin{array}{cccc|cccc} \sqrt{\frac{k}{n}} & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & \sqrt{\frac{k}{n}} & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \cdots & & & & \cdots & & & \\ 0 & 0 & \cdots & \sqrt{\frac{k}{n}} & 0 & 0 & \cdots & 0 \\ \hline 0 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & 1 & \cdots & 0 \\ \cdots & & & & \cdots & & & \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 1 \end{array}\right],$$

an $n \times n$ matrix whose columns are n mutually orthogonal vectors of the needed lengths.

Let

$$C = \left[\begin{array}{ccc|cccc} \frac{1}{\sqrt{k}} & \cdots & \frac{1}{\sqrt{k}} & 0 & & \cdots & 0 \\ \cdots & \cdots & & \cdots & & & \cdots \\ \cdots & \cdots & & 0 & & \cdots & 0 \\ \hline 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & 1 & \cdots & 0 \\ \cdots & & & & & & \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 1 \end{array}\right],$$

an $n \times n$ orthogonal matrix which rotates the first k columns of D into vectors whose endpoints lie on the plane $a_1 = n^{-\frac{1}{2}}$ and which leaves the last $n - k$ columns unchanged.

Let B be an $n \times n$ orthogonal matrix whose first column consists entirely of the elements $n^{-\frac{1}{2}}$. Since B rotates the plane $a_1 = n^{-\frac{1}{2}}$ into the plane $\sum_{i=1}^{n} a_i = 1$, it must also rotate the parallel plane $a_1 = 0$ into $\sum_{i=1}^{n} a_i = 0$.

Thus $A = BCD$ is an $n \times n$ matrix whose columns are orthogonal vectors. The first k are of length $(k/n)^{\frac{1}{2}}$ and have endpoints on $\sum_{i=1}^{n} a_i = 1$; the last $n - k$ are of length one and have endpoints on $\sum_{i=1}^{n} a_i = 0$.

Then let

$$t_i = \frac{z_i - \mu_i}{\sqrt{\frac{k}{n(n-k)} \sum_{m=1}^{n-k} u_m^2}}, \qquad i = 1, \cdots, k.$$

These are k t-variables whose numerators are independent but whose denominators are the same. Their frequency function is (see [2]):

$$f_{n-k}(t_1, \cdots, t_k) = \frac{\Gamma\left(\frac{n}{2}\right)}{[\pi(n-k)]^{k/2}\Gamma\left(\frac{n-k}{2}\right)} \left(1 + \frac{\sum_{i=1}^{k} t_i^2}{n-k}\right)^{-(n/2)}$$

If c_α is defined by

$$\int_{-c_\alpha}^{c_\alpha} \cdots \int_{-c_\alpha}^{c_\alpha} f_{n-k}(t_1, \cdots, t_k)\, dt_1, \cdots, dt_k = 1 - \alpha,$$

then $P(-c_\alpha < t_1 < c_\alpha, \cdots, -c_\alpha < t_k < c_\alpha) = 1 - \alpha$. Thus an exact set of confidence intervals of equal but variable lengths is obtained:

$$z_i \pm c_\alpha \sqrt{\frac{k}{n(n-k)} \sum_{m=1}^{n-k} u_m^2}, \qquad i = 1, \cdots, k.$$

4. Intervals of bounded confidence level using the chi-square distribution and Hotelling's T-distribution.

4.1. Known variances. For a sample of size n from the multivariate normal distribution with means $\mu_1, \cdots, \mu_k$ and covariance matrix (λ_{is}), the expression

$n\sum_{i=1}^{k}\sum_{s=1}^{k}\lambda^{is}(\bar{y}_i - \mu_i)(\bar{y}_s - \mu_s)$ follows a Chi-square distribution with k degrees of freedom. Here λ^{is} denotes an element of the inverse matrix $(\lambda^{is}) = (\lambda_{is})^{-1}$, and $\bar{y}_i$ is the sample mean of the observations on y_i. Then

$$\sum_{i=1}^{k}\sum_{s=1}^{k}\lambda^{is}(\bar{y}_i - \mu_i)(\bar{y}_s - \mu_s) = \frac{c_\alpha}{n},$$

where c_α is defined by $U_k(c_\alpha) = 1 - \alpha$, with U_k the cumulative distribution function of a Chi-square variable with k degrees of freedom. In the parameter space of the $\mu_1, \cdots, \mu_k$, this equation defines an ellipsoid, which will be denoted by E. Then

$$P\left(\sum_{i=1}^{k}\sum_{s=1}^{k}\lambda^{is}(\bar{y}_i - \mu_i)(\bar{y}_s - \mu_s) < \frac{c_\alpha}{n}\right) = P[E \text{ covers } (\mu_1, \cdots, \mu_k)] = 1 - \alpha.$$

To obtain a rectangular confidence region of bounded confidence level, a rectangular parallelepiped, say R, with boundary planes parallel to the coordinate planes in the $\mu_1, \cdots, \mu_k$ space is circumscribed around the ellipsoid E. The boundary planes of R are found to be

$$\mu_i = \bar{y}_i \pm \frac{\sigma_i}{\sqrt{n}}\sqrt{c_\alpha},$$

and are not dependent on the correlations.

Then $P[R \text{ covers } (\mu_1, \cdots, \mu_k)] > P[E \text{ covers } (\mu_1, \cdots, \mu_k)] = 1 - \alpha$, thus giving a set of intervals, $\bar{y}_i \pm (\sigma_i/n^{\frac{1}{2}})c_\alpha^{\frac{1}{2}}$, with $U_k(c_\alpha) = 1 - \alpha$.

4.2. Unknown variances. The same method applies when the variances are unknown and $n > k$, using Hotelling's T-statistic. Here E is the ellipsoid $\sum_{i=1}^{k}\sum_{s=1}^{k} l^{is}(\bar{y}_i - \mu_i)(\bar{y}_s - \mu_s) = c_\alpha^2/n$ where (l^{is}) is the inverse of the matrix (l_{is}) and $l_{is} = \sum_{j=1}^{n}(y_{ij} - \bar{y}_i)(y_{sj} - \bar{y}_s)/(n-1)$, $i = 1, \cdots, k$; $s = 1, \cdots, k$. The boundary planes of R, the circumscribed parallelepiped, are $\mu_i = \bar{y}_i \pm (\hat{\sigma}_i/n^{\frac{1}{2}})c_\alpha$, where $\hat{\sigma}_i = l_{ii}^{\frac{1}{2}}$. For c_α defined by $F(c_\alpha) = 1 - \alpha$, with F the c.d.f. of Hotelling's T, the set of confidence intervals is $\bar{y}_i \pm (\hat{\sigma}_i/n^{\frac{1}{2}})c_\alpha$, $i = 1, 2, \cdots, k$.

It is to be noted that this is the only set of intervals given ln this paper for which no assumption has been made concerning the variances. For the other sets, the variances were assumed to be known or else to be unknown but equal.

4.3. More general distribution functions. For n large, T^2 can be assumed to follow a Chi-square distribution with k degrees of freedom, even though the original variables are not normally distributed [3]. A set of confidence intervals for $\mu_1, \cdots, \mu_k$ is then $\bar{y}_i \pm (\hat{\sigma}_i/n^{\frac{1}{2}})c_\alpha^{\frac{1}{2}}$, with c_α the upper α point of the Chi-square distribution with k degrees of freedom.

5. Bounded regions based on linear contrasts. Henry Scheffé [4] obtains simultaneous confidence intervals for the totality of linear contrasts among k means, $\mu_1, \cdots, \mu_k$, using the F distribution. He shows that $P(\hat{\theta} - S\hat{\sigma}_{\hat{\theta}} \leqq \theta \leqq \theta + S\hat{\sigma}_{\hat{\theta}}) = 1 - \alpha$. Here θ is any linear contrast; $S^2 = (k-1)c_\alpha$; c_α is the upper α point of the F distribution with $k - 1$ and ν degrees of freedom; ν is the de-

grees of freedom of the χ^2 variable used in estimating the variance; and P denotes the probability that all such intervals cover their corresponding contrasts.

It can easily be shown that confidence intervals for the totality of linear combinations of $\mu_1, \cdots, \mu_k$ are similarly obtained from $P(\theta - S\hat{\sigma}_{\hat{\theta}} \leq \hat{\theta} \leq \theta + S)\hat{\sigma}_{\hat{\theta}} = 1 - \alpha$, where now $S^2 = kc_\alpha$, with c_α the upper α point of the F distribution with k and ν degrees of freedom. Since the k means $\mu_1, \cdots, \mu_k$ are a subset of the linear combinations, confidence intervals for them follow immediately.

5.1. Variances known. If the variables $y_1, \cdots, y_k$ are normally distributed with unknown means $\mu_1, \cdots, \mu_k$, known variances $\sigma_1^2, \cdots, \sigma_k^2$, and unknown correlations, ρ_{is}, then the χ^2 distribution is used rather than the F distribution, and we have:

$$P\left(\bar{y}_1 - \frac{\sigma_1}{\sqrt{n}}\sqrt{c_\alpha} < \mu_1 < \bar{y}_1 + \frac{\sigma_1}{\sqrt{n}}\sqrt{c_\alpha}, \cdots, \bar{y}_k - \frac{\sigma_k}{\sqrt{n}}\sqrt{c_\alpha} < \mu_k < \bar{y}_k + \frac{\sigma_k}{\sqrt{n}}\sqrt{c_\alpha}\right) \geqq 1 - \alpha.$$

Here c_α is, as in section 4.1, the upper α point of the χ^2 distribution with k degrees of freedom, and the intervals obtained are the same as those of section 4.1.

5.2. Variances unknown but equal. When the variances are unknown but equal, then as an estimate of σ^2 one may use $\hat{\sigma}_1^2 = \sum_{j=1}^n (y_{1j} - \bar{y}_1)^2/(n-1)$. Then

$$P\left(\bar{y}_1 - \sqrt{\frac{k}{n}}\hat{\sigma}_1\sqrt{c_\alpha} < \mu_1 < \bar{y}_1 + \sqrt{\frac{k}{n}}\hat{\sigma}_1\sqrt{c_\alpha}, \cdots, \bar{y}_k - \sqrt{\frac{k}{n}}\hat{\sigma}_1\sqrt{c_\alpha} < \mu_k < \bar{y}_k + \sqrt{\frac{k}{n}}\hat{\sigma}_1\sqrt{c_\alpha}\right) \geqq 1 - \alpha,$$

with c_α the upper α point of the F distribution with k and $n - 1$ degrees of freedom. The confidence intervals are $\bar{y}_i \pm (k/n)^{\frac{1}{2}}\hat{\sigma}_1 c_\alpha^{\frac{1}{2}}$.

It may seem unsatisfactory to use only the data from one sample point as an estimate of σ^2; this has been done in order to have a χ^2 variable for the denominator of the F variable.

If one wishes to use a pooled estimate of the variance, $\hat{\sigma}_p^2 = \sum_{i=1}^k \hat{\sigma}_i^2/k$, then $\hat{\sigma}_p^2$ no longer has a χ^2 distribution because of the dependence of the variables. It is possible to show, however, that the F distribution may still be used, provided for degrees of freedom one uses k and $n - 1$ (rather than k and $k(n-1)$). That the degrees of freedom may not be increased may be seen by examining the extreme case when all the correlations are equal to one.

To establish the necessary inequality for using $\hat{\sigma}_p^2$, one may fix $\hat{\sigma}_1, \cdots, \hat{\sigma}_k$ and consider the conditional probability

$$P\left(\bar{y}_1 - \sqrt{\frac{k}{n}}\hat{\sigma}_p\sqrt{c_\alpha} < \mu_1 < \bar{y}_1 + \sqrt{\frac{k}{n}}\hat{\sigma}_p\sqrt{c_\alpha}, \cdots, \bar{y}_k - \sqrt{\frac{k}{n}}\hat{\sigma}_p\sqrt{c_\alpha} < \mu_k < \bar{y}_1 + \sqrt{\frac{k}{n}}\hat{\sigma}_p\sqrt{c_\alpha} \,\middle|\, \hat{\sigma}_1, \cdots, \hat{\sigma}_k\right)$$

$$\geqq P\left(\bar{y}_1 - \sqrt{\frac{k}{n}}\frac{\Sigma\hat{\sigma}_i}{k}\sqrt{c_\alpha} < \mu_1 < \bar{y}_1 + \sqrt{\frac{k}{n}}\frac{\Sigma\hat{\sigma}_i}{k}\sqrt{c_\alpha}, \cdots, \bar{y}_k - \sqrt{\frac{k}{n}}\frac{\Sigma\hat{\sigma}_i}{k}\sqrt{c_\alpha} < \mu_k < \bar{y}_k + \sqrt{\frac{k}{n}}\frac{\Sigma\hat{\sigma}_i}{k}\sqrt{c_\alpha} \mid \hat{\sigma}_1, \cdots, \hat{\sigma}_k\right)$$

$$\geqq \sum_{i=1}^{k} P\left(\bar{y}_1 - \sqrt{\frac{k}{n}}\hat{\sigma}_i\sqrt{c_\alpha} < \mu_1 < \bar{y}_1 + \sqrt{\frac{k}{n}}\hat{\sigma}_i\sqrt{c_\alpha}, \cdots, \bar{y}_k - \sqrt{\frac{k}{n}}\hat{\sigma}_i\sqrt{c_\alpha} < \mu_k < \bar{y}_k + \sqrt{\frac{k}{n}}\hat{\sigma}_i\sqrt{c_\alpha} \mid \hat{\sigma}_1, \cdots, \hat{\sigma}_k\right)\Big/ k$$

Thus for the unconditional probability one has:

$$P\left(\bar{y}_1 - \sqrt{\frac{k}{n}}\hat{\sigma}_p\sqrt{c_\alpha} < \mu_1 < \bar{y}_1 + \sqrt{\frac{k}{n}}\hat{\sigma}_p\sqrt{c_\alpha}, \cdots, \bar{y}_k - \sqrt{\frac{k}{n}}\hat{\sigma}_p\sqrt{c_\alpha} < \mu_k < \bar{y}_k + \sqrt{\frac{k}{n}}\hat{\sigma}_p\sqrt{c_\alpha}\right) \geqq 1 - \alpha.$$

6. Regions based on a bonferroni inequality. Confidence regions can be obtained very simply using a Bonferroni inequality [5]. The use of this inequality in a related situation was suggested by E. Paulson [6].

6.1. Variances known. Let $n_k(y_1, \cdots, y_k; \cdot\mu_i, \sigma_i^2, \rho_{is})$ be the frequency function of k normally distributed variables with means $\mu_1, \cdots, \mu_k$, known variances $\sigma_1^2, \cdots, \sigma_k^2$, and unknown correlations ρ_{is}. Let $\bar{y}_i$ be the mean of a random sample of size n, $y_{i1}, \cdots, y_{in}$.

Let $z_i = ((\bar{y}_i - \mu_i)n^{\frac{1}{2}})/\sigma_i)$, $i = 1, \cdots, k$. Then the joint frequency function of $z_1, \cdots, z_k$ is $n_k(z_1, \cdots, z_k; 0, 1, \rho_{is})$, and

$$P(-c < z_1 < c, \cdots, -c < z_k < c) = \int_{-c}^{c} \cdots \int_{-c}^{c} n_k(z_1, \cdots, z_k; 0, 1, \rho_{is})\, dz_1 \cdots dz_k.$$

Using a Bonferroni inequality, this integral is greater than or equal to $1 - 2k(1 - N(c))$, where N is the c.d.f. of a standard normal variable. Setting this expression equal to $1 - \alpha$, c_α may be defined by $N(c_\alpha) = 1 - (\alpha/2k)$. Then

$$P\left(-c_\alpha < \frac{(\bar{y}_1 - \mu_1)\sqrt{n}}{\sigma_1} < c_\alpha, \cdots, -c_\alpha < \frac{(\bar{y}_k - \mu_k)\sqrt{n}}{\sigma_k} < c_\alpha\right) = P[R \text{ covers } (\mu_1, \cdots, \mu_k)] \geqq 1 - \alpha,$$

where R is bounded by

$$\mu_1 = \bar{y}_i \pm \frac{\sigma_i}{\sqrt{n}} c_\alpha.$$

6.2. Variances unknown but equal. Let y_i, $i = 1, \cdots, k$, have the joint frequency function $n_k(y_1, \cdots, y_k; \mu_i, \sigma^2, \rho_{is})$, where the variances are unknown but equal. Let $z_i = ((\bar{y}_i - \mu_i)n^{\frac{1}{2}})/\sigma$, $i = 1, \cdots, k$.

We wish to define Student t-variables $t_1, \cdots, t_k$ using $z_1, \cdots, z_k$ in the numerators and using the same Chi-square variable in the denominators. If $u_i = \sum_{j=1}^{n} (y_{ij} - \bar{y}_i)^2/\sigma^2$, then u_i is a Chi-square variable with $n - 1$ degrees of freedom. Since the u_i are not independent of each other, we choose one, say u_1, to use in all the denominators, rather than use their sum which does not have a Chi-square distribution.

Then

$$t_i = \frac{\sqrt{n-1}\, z_i}{u_1^{1/2}} = \frac{\sqrt{n}(\bar{y}_i - \mu_i)}{\hat{\sigma}_1}, \qquad i = 1, \cdots, k,$$

are Student t-variables with the same denominators. Their distribution function [3] is;

$$f_{n-1}(t_1, \cdots, t_k; \rho_{is}) = \frac{\Gamma\left(\frac{k+n-1}{2}\right)}{(n-1)^{k/2}\pi^{k/2}\Gamma\left(\frac{n-1}{2}\right)}$$
$$|(\rho^{is})|^{1/2}\left[1 + (n-1)^{-1}\sum_{i=1}^{k}\sum_{s=1}^{k}\rho^{is}t_it_s\right] - \frac{k+n-1}{2}$$

where ρ^{is} is an element of $(\rho^{is}) = (\rho_{is})^{-1}$, and $|(\rho^{is})|$ is the determinant of (ρ^{is}).

As in 6.1,

$$P(-c < t_1 < c, \cdots, -c < t_k < c)$$
$$= \int_{-c}^{c}\cdots\int_{-c}^{c} f_{n-1}(t_1, \cdots, t_k; \rho_{is})\, dt_1 \cdots dt_k \geq 1 - 2k(1 - H_{n-1}(c)),$$

where H_{n-1} is the c.d.f. of a t-variable with $n - 1$ degrees of freedom.

The set of confidence intervals is then

$$\bar{y}_i \pm \frac{\hat{\sigma}_1}{\sqrt{n}} c_\alpha,$$

where

$$H_{n-1}(c_\alpha) = 1 - \frac{\alpha}{2k},$$

and

$$\hat{\sigma}_1 = \sum_{j=1}^{n} (y_{1j} - \bar{y}_1)^2/(n-1).$$

As in section 5.2, it is possible in these confidence intervals to replace $\hat{\sigma}_1$ by $\hat{\sigma}_p$, the pooled estimate of the variance; $n - 1$ must be retained as the degrees of freedom.

7. Regions with bounded confidence level using inequalities between dependent and independent cases.

7.1. Variances known. For $y_1, \cdots, y_k$ independently normally distributed with unknown means $\mu_1, \cdots, \mu_k$ and known variances, $\sigma_1^2, \cdots, \sigma_k^2$, let x_i be

defined by $x_i = (n^{\frac{1}{2}}(\bar{y}_i - \mu_i))/\sigma_i$, where $\bar{y}_i$ is the mean of the n observations on the ith variable. Then

$$P(-c_\alpha < x_1 < c_\alpha, \cdots, -c_\alpha < x_k < c_\alpha) = \prod_{i=1}^{k} P(-c_\alpha < x_i < c_\alpha) = 1 - \alpha,$$

where c_α is defined by $N(c_\alpha) = \frac{1}{2}[1 + (1 - \alpha)^{1/k}]$, with N the c.d.f. of the univariate normal distribution. The set of simultaneous confidence intervals whose exact confidence level is $1 - \alpha$ is then $\bar{y}_i \pm \sigma_i c_\alpha / n^{\frac{1}{2}}$.

If, now, the $y_1, \cdots, y_k$ are defined as above except that now there may be correlations among them, the same confidence intervals can be used as a set with bounded confidence level, provided it can be proved that

$$P(-c_\alpha < x_1 < c_\alpha, \cdots, -c_\alpha < x_k < c_\alpha) \geqq 1 - \alpha.$$

The proof of the following theorem establishes this inequality for certain cases.

THEOREM. If $x_1, \cdots, x_k$ are normally distributed with zero means, unit variances, and correlations ρ_{is}, then

$$\int \underset{C}{\cdots} \int n_k(x_1, \cdots, x_k; 0, 1, \rho_{is})\, dx_1 \cdots dx_k \geqq \left[\int_{x=-c}^{x=c} n_1(x; 0, 1)\, dx\right]^k,$$

provided (1) $k = 2$ or 3; or (2) $\rho_{is} = b_i b_s$, for $i, s = 1, 2, \cdots, k, i \neq s$ and with $0 < b_i < 1, i = 1, 2, \cdots, k$. The region of integration C is the region bounded by the planes $x_i = \pm c, i = 1, \cdots, k$; $n_k(x_1, \cdots, x_k; 0, 1, \rho_{is})$ is the frequency function of $x_1, \cdots, x_k$; and $n_1(x; 0, 1)$ is the standard univariate normal frequency function.

PROOF. (1) $k = 2, 3$. For brevity the proof is merely outlined. The expression $\int \underset{C}{\cdots} \int n_k(x_1, \cdots, x_k; 0, 1, \rho_{is})\, dx_1, \cdots, dx_k$ may be regarded as a function of the ρ_{is}, say $F(\rho_{is})$. The proof consists in showing that for all admissible ρ_{is}, $F(\rho_{is})$ has an absolute minimum at the origin of the ρ_{is} space.

First it must be shown that there is a relative minimum at the origin. This can be shown for any k by considering the various first and second partial derivatives with respect to the correlations.

The first partial derivative with respect to ρ_{12}, say F_{12}, can be shown to be:

$$F_{12} = 2\int_{x_3=-c}^{x_3=c} \cdots \int_{x_k=-c}^{x_k=c}$$
$$\cdot [n_k(c, c, x_3, \cdots, x_k; 0, 1, \rho_{is}) - n_k(c, -c, x_3, \cdots, x_k; 0, 1, \rho_{is})]\, dx_3, \cdots, dx_k \cdot$$

Similarly, the second derivative with respect to ρ_{12} and ρ_{pq}, say $F_{12.pq}$, is

$$F_{12.pq} = 2\int_{x_3=-c}^{x_3=c} \cdots \int_{x_k=-c}^{x_k=c} n_k(c, c, x_3, \cdots, x_k; 0, 1, \rho_{is})$$
$$\cdot \left[\rho^{pq} + \left(\sum_{\substack{i=1\\x_1=c\\x_2=c}}^{k} \rho^{pi} x_i\right)\left(\sum_{\substack{i=1\\x_1=c\\x_2=c}}^{k} \rho^{qi} x_i\right)\right]$$
$$dx_3, \cdots, dx_k \text{—a similar integral with } x_1 = c, x_2 = -c.$$

When all the ρ_{is}'s are zero, it is easily seen that F_{12} vanishes. Further, $F_{12,pq}$ vanishes also at the origin unless $p = 1$ and $q = 2$, while $F_{12,12}$ is seen to be positive.

Thus in the expansion of $F(\rho_{is})$ about the origin, the first degree terms vanish and the second degree terms form a positive definite quadratic form, so that $F(\rho_{is})$ has a relative minimum at the origin for any k.

The next part of the proof is to show from the form of the first derivative, that at any point beside the origin, at least one of the first derivatives differ from zero. This was done only for $k = 2$ and 3.

The set of all admissible points (points such that (ρ_{is}) is positive definite and $0 < |(\rho_{is})| < 1$), together with the boundary points, form a compact set, so that $F(\rho_{is})$ must assume an absolute minimum either at an admissible point or at a boundary point. Hence if it can be shown that no point on the boundary of the set yields an absolute minimum, then the absolute minimum of F must be at the origin.

For $k = 2$, the boundary points are just $\rho_{12} = \pm 1$, and they actually yield absolute maxima for $F(\rho_{12})$.

For $k = 3$, a boundary point, say $(\rho_{12}, \rho_{13}, \rho_{23})$ was considered. It was shown that for m sufficiently close to 1 but less than 1, $(m\rho_{12}, m\rho_{13}, \rho_{23})$ is an admissible point, and that the derivative of F at $(m\rho_{12}, m\rho_{13}, \rho_{23})$ in the direction of $(\rho_{12}, \rho_{13}, \rho_{23})$ is positive. Hence $(\rho_{12}, \rho_{13}, \rho_{23})$ cannot yield an absolute minimum of F.

This completes the outline of the proof for $k = 2$ and 3, with any correlation matrix.

(2) For any k, if $\rho_{is} = b_i b_s$, with $0 < b_i < 1$ for $i = 1, \cdots, k$, a proof may be given which is adapted from the proof of a similar theorem by C. W. Dunnett and M. Sobel [3].

For $y_0, y_1, \cdots, y_k$ independently normally distributed, with zero means and unit variances, define

$$x_i = \sqrt{1 - b_i^2}\, y_i - b_i y_0, \qquad i = 1, \cdots, k.$$

Then the x_i's are normally distributed with means zero, unit variances, and correlations $\rho_{is} = b_i b_s$.

The theorem may be restated as follows:

$$P(-c < x_1 < c, \cdots, -c < x_k < c) \geq \prod_{i=1}^{k} P(-c < x_i < c),$$

or

$$P(-c < \sqrt{1 - b_1^2}\, y_1 - b_1 y_0 < c, \cdots, -c < \sqrt{1 - b_k^2}\, y_k - b_k y_0 < c)$$
$$\geq \prod_{i=1}^{k} P(-c < \sqrt{1 - b_i^2}\, y_i - b_i y_0 < c).$$

or

$$P(d_1 < y_1 < e_1, \cdots, d_k < y_k < e_k) \geq \prod_{i=1}^{k} P(d_i < y_i < e_i),$$

where

$$d_i = \frac{-c + b_i y_0}{\sqrt{1 - b_i^2}}, \qquad e_i = \frac{c + b_i y_0}{\sqrt{1 - b_i^2}}, \qquad i = 1, \cdots, k.$$

This may be written as:

$$\int_{y_0=-\infty}^{y_0=\infty} \left[\int_{y_1=d_1}^{y_1=e_1} \cdots \int_{y_k=d_k}^{y_k=e_k} n_k(y_1, \cdots, y_k; 0, 1, 0)\, dy_1, \cdots, dy_k\right] n_1(y_0; 0, 1)\, dy_0$$
$$\geqq \prod_{i=1}^{k} \int_{y_0=-\infty}^{y_0=\infty} \left[\int_{y_i=d_i}^{y_i=e_i} n_1(y_i; 0, 1)\, dy_i\right] n_1(y_0; 0, 1)\, dy_0,$$

or

$$\int_{y_0=-\infty}^{y_0=\infty} \left[\prod_{i=1}^{k} F_i(y_0)\right] n_1(y_0; 0, 1)\, dy_0 \geqq \prod_{i=1}^{k} \int_{y_0=-\infty}^{y_0=\infty} F_i(y_0) n_1(y_0; 0, 1)\, dy_0,$$

where

$$F_i(y_0) = \int_{d_i}^{e_i} n_1(y_i; 0, 1)\, dy_i.$$

Thus the inequality becomes:

$$E\left(\prod_{i=1}^{k} F_i(y_0)\right) \geqq \prod_{i=1}^{k} E(F_i(y_0)).$$

The expected value of a product of monotone bounded functions is greater than or equal to the product of their expected values [6], so that the last inequality would hold if the F_i were monotone. The functions $F_i(y_0)$, however, are seen to increase from $-\infty$ to 0 and to decrease from 0 to ∞. Since the frequency function of y_0 is symmetric about the origin, the transformation $z = |y_0|$ changes the inequality to

$$E\left(\prod_{i=1}^{k} F_i(z)\right) \geqq \prod_{i=1}^{k} E(F_i(z)),$$

where $F_i(z)$ are monotonically decreasing bounded functions. This completes the proof of the theorem.

7.2. Variances unknown but equal. When the variances are unknown but equal, Student t-variables t_i with the joint frequency function

$$f_{n-1}(t_1, \cdots, t_k; \rho_{is}),$$

as defined in 6.2, are used to form confidence intervals. Using the same methods as in 7.1, the following theorem can be proved:

THEOREM. *For $k = 2$ or 3,*

$$\int \underset{C}{\cdots} \int f_{n-1}(t_1, \cdots, t_k; \rho_{is})\, dt_1 \cdots dt_k \geqq \int \underset{C}{\cdots} \int f_{n-1}(t, \cdots, t_k; 0)\, dt_1 \cdots dt_k$$

For $\rho_{is} = b_i b_s$, with $0 < b_i < 1$, $i = 1, \cdots, k$,

$$\int \underset{C}{\cdots} \int f_{n-1}(t_1, \cdots, t_k ; \rho_{is})\, dt_1 \cdots dt_k \geqq \left[\int_{-c}^{c} f_{n-1}(t)\, dt\right]^k.$$

In this theorem, whose proof follows the same lines as the one in 7.1, C is the region bounded by $t_i = \pm c$, $i = 1, \cdots, k$, $f_{n-1}(t)$ is the density function of a Student t-variable with $n - 1$ degrees of freedom, and $f_{n-1}(t_1, \cdots, t_k ; 0)$ is the joint frequency function of the t-variables when all ρ_{is} are zero.

Since

$$t_i = \frac{\sqrt{n}(\bar{y}_i - \mu_i)}{\sigma_1}, \qquad i = 1, \cdots, k,$$

sets of confidence intervals obtained are as follows:

For $k = 2$ or 3, $\bar{y}_1 \pm (\hat{\sigma}_1/n^{\frac{1}{2}})c_\alpha$, where c_α is defined by

$$\int_{-c_\alpha}^{c_\alpha} \cdots \int_{-c_\alpha}^{c_\alpha} f_{n-1}(t_1, \cdots, t_k ; 0)\, dt_1 \cdots dt_k = 1 - \alpha.$$

For any k and $\rho_{is} = b_i b_s$, $0 < b_i < 1$, $i = 1, \cdots, k$, the same set is obtained, but with c_α defined by $H_{n-1}(c_\alpha) = (1 + (1 - \alpha)^{1/k})/2$, where H_{n-1} is the c.d.f. of a Student t-variable with $n - 1$ degrees of freedom. As in sections 5.2 and 6.2 one may use $\hat{\sigma}_p^2$ in place of $\hat{\sigma}_1^2$, provided one keeps $n - 1$ as the degrees of freedom.

8. Comparison of confidence intervals. In Table I are listed various sets of confidence intervals, with their properties and restrictions.

One rather obvious way to compare them is by comparing their lengths, or the expected values of the lengths. In Table II are given numerical values of d_α for $1 - \alpha = .95$, where

$$d_\alpha = \frac{\sqrt{n}}{\sigma} \sqrt{E(\tfrac{1}{2}\ell)^2},$$

with ℓ the length of the confidence interval. Throughout Table II, the variances are assumed to be equal.

When the variances are known and equal, and all the correlations are zero, the shortest set of confidence intervals must be those of section 7.1. When nothing is known about the correlations, no shorter set can be obtained. The last column in section 7 of Table II therefore gives the smallest obtainable values for d_α, and may be used as a standard for comparison.

For $1 - \alpha = .95$, the Bonferroni inequality intervals of section 6 are almost as good as the best ones. Indeed for $1 - \alpha$ as low as .80, the values of d_α are still very close, being:

k	Bonferroni	"Best"
1	1.28	1.28
2	1.64	1.61
4	1.96	1.92
6	2.13	2.09
8	2.24	2.20
10	2.33	2.29

TABLE I

Confidence Intervals for Means of Dependent, Normally Distributed Variables

Section	Confidence Intervals	Definition of c_α	Conditions
3.1	$\sum_{j=1}^{n} a_{ji} y_{ij} \pm \frac{k}{n} \sigma_i \cdot c_\alpha$	$N(c_\alpha) = \frac{1 + (1-\alpha)^{1/k}}{2}$	$n \geqq k$ (1)
3.2	$\sum_{j=1}^{n} a_{ji} y_{ij} \pm \sqrt{\frac{k}{n(n-k)} \sum_{m=1}^{n-k} u_m^2} \cdot c_\alpha$	$H_{n-k}(c_\alpha) = \frac{1 + (1-\alpha)^{1/k}}{2}$	$n > k$ (2, 3)
4.1	$\bar{y}_i \pm \frac{\sigma_i}{\sqrt{n}} \cdot \sqrt{c_\alpha}$	$U_k(c_\alpha) = 1 - \alpha$	(4)
4.2	$\bar{y}_i \pm \frac{\hat{\sigma}_i}{\sqrt{n}} \cdot c_\alpha$	$F(c_\alpha) = 1 - \alpha$	$n > k$ (5)
5.1	$\bar{y}_i \pm \frac{\sigma_i}{\sqrt{n}} \sqrt{c_\alpha}$	$U_k(c_\alpha) = 1 - \alpha$	(4)
5.2	$\bar{y}_i \pm \frac{\hat{\sigma}_1}{\sqrt{n}} \sqrt{c_\alpha}$	$F_{k,n-1}(c_\alpha) = 1 - \alpha$	(6)
6.1	$\bar{y}_i \pm \frac{\sigma_i}{\sqrt{n}} \cdot c_\alpha$	$N(c_\alpha) = 1 - \frac{\alpha}{2k}$	(1)
6.2	$\bar{y}_i \pm \frac{\hat{\sigma}_1}{\sqrt{n}} \cdot c_\alpha$	$H_{n-1}(c_\alpha) = 1 - \frac{\alpha}{2k}$	(2)
7.1	$\bar{y}_i \pm \frac{\sigma_i}{\sqrt{n}} \cdot c_\alpha$	$N(c_\alpha) = \frac{1 + (1-\alpha)^{1/k}}{2}$ $k = 2, 3,$ or $\rho_{is} = b_i b_s$	(1)
7.2	$\bar{y}_i \pm \frac{\hat{\sigma}_1}{\sqrt{n}} \cdot c_\alpha$	$H_{n-1}(c_\alpha) = \frac{1 + (1-\alpha)^{1/k}}{2}$ $k = 2, 3,$ or $\rho_{is} = b_i b_s$	(2, 3) (2)

(1) N is the cumulative standard normal distribution function.

(2) H_ν is the cumulative distribution function of a Student t-variable with ν degrees of freedom.

(3) This definition of c_α is approximate. The exact definition is:

$$\int_{-c_\alpha}^{c_\alpha} \cdots \int_{-c_\alpha}^{c_\alpha} f_\nu(t_1, \cdots, t_k)\, dt_1, \cdots, dt_k = 1 - \alpha, \quad \text{where} \quad f_\nu(t_1, \cdots, t_k)$$

$$= \frac{\Gamma\left(\frac{k+\nu}{2}\right)}{\nu^{k/2}\pi^{k/2}\Gamma\left(\frac{\nu}{2}\right)} \left[1 + \frac{\sum_{i=1}^{k} t_i^2}{\nu}\right]^{-(k+\nu)/2}, \quad \text{where}$$

ν is the degrees of freedom of t_i.

(4) U_k is the cumulative distribution function of a Chi-square variable with k degrees of freedom.

(5) F is the cumulative distribution function of Hotelling's T.

(6) $F_{k,n-1}$ is the cumulative distribution function of an F variable with k and $n - 1$ degrees of freedom.

TABLE II

*Comparison of Lengths of Confidence Intervals for Means of Dependent, Normally Distributed Variables with Equal Variances, $1 - \alpha = .95$**

k	n							
	Variances Unknown						Variances Known	
	Section	4	6	8	10	20	Section	Any
	4.1						4.2	
1		3.18	2.57	2.36	2.26	2.09		1.96
2		10.8	5.52	4.53	4.12	3.55		3.17
4			27.9	11.4	8.63	6.13		4.98
6				40.3	17.8	8.75		6.44
8					77.0	11.8		7.72
10						15.6		8.85
	5.1						5.2	
1		3.18	2.57	2.36	2.26	2.09		1.96
2		7.55	4.16	3.47	3.17	2.74		2.45
4			13.9	6.70	5.21	3.79		3.08
6				20.1	9.12	4.82		3.55
8					26.4	6.01		3.94
10						7.53		4.28
	6.1						6.2	
1		3.18	2.57	2.36	2.26	2.09		1.96
2		4.37	3.40	3.08	2.92	2.66		2.45
4		6.04	4.56	4.06	3.81	3.41		3.08
6		7.32	5.45	4.82	4.50	3.98		3.55
8		8.41	6.21	5.37	5.08	4.46		3.94
10		9.38	6.88	6.03	5.60	4.89		4.28
	7.1						7.2	
1		3.18	2.57	2.36	2.26	2.09		1.96
2		4.17	3.16	2.84	2.68	2.44		2.24
4		5.41	3.80	3.33	3.11	2.76		2.50
6		6.22	4.22	3.64	3.36	2.94		2.64
8		6.92	4.53	3.86	3.55	3.07		2.74
10		7.47	4.77	4.03	3.69	3.17		2.81
	8.1						8.2	
1		3.18	2.57	2.36	2.26	2.09		1.96
2		4.16	3.15	2.83	2.68	2.43		2.24
4		5.35	3.79	3.32	3.10	2.75		2.49
6		6.17	4.20	3.62	3.35	2.94		2.63
8		6.86	4.50	3.84	3.53	3.07		2.73
10		7.40	4.76	4.01	3.67	3.16		2.80

* The figures given in the table are values of $(n^{\frac{1}{2}}/\sigma)\sqrt{E(\frac{1}{2}\ell)^2}$, where ℓ is the length of the confidence interval.

It would be interesting to show that the "best" intervals can be used for arbitrary k and arbitrary correlations, but from a practical viewpoint, for $1 - \alpha$ large enough to be of interest, the Bonferroni regions are good enough.

The regions of section 5, based on the T-distribution and the χ^2 distribution, compare favorably only when k is small and n relatively large. The regions with exact confidence level are everywhere unnecessarily long.

REFERENCES

[1] PAUL G. HOEL, "Confidence bands for polynomial curves," *Ann. Math. Stat.*, Vol. 25, (1954), pp. 534–542.

[2] C. W. DUNNET AND M. SOBEL, "Approximations to the probability integral and certain percentage points of a multivariate analogue of Student's t-distribution," *Biometrika*, Vol. 42, (1955), pp. 258–260.

[3] H. B. MANN AND A. WALD, "On stochastic limit and order relationships," *Ann. Math. Stat.*, Vol. 14, (1943), p. 224.

[4] HENRY SCHEFFÉ, "A method of judging all contrasts in the analysis of variance," *Biometrika*, Vol. 40, (1953), pp. 87–104.

[5] W. FELLER, *An introduction to probability theory and its applications.* I, p. 75, John Wiley and Sons, 1950.

[6] E. PAULSON, "On the comparison of several experimental categories with a control," *Ann. Math. Stat.*, Vol. 23, (1952), p. 249.

[7] A. W. KIMBALL, "On dependent tests of significance in the analysis of variance," *Ann. Math. Stat.*, Vol. 22, (1951), pp. 600–602.

A MARKOVIAN FUNCTION OF A MARKOV CHAIN[1]

By C. J. Burke and M. Rosenblatt

Indiana University

1. Statement of the problem and the results obtained. Consider a Markov chain $X(n), n = 0, 1, 2, \cdots$, with a finite number of states $1, \cdots, m$ and stationary transition probability matrix $P = (p_{ij})$

$$(1) \qquad \begin{aligned} p_{ij} = P[X(n+1) = j \mid X(n) = i] \geqq 0, \qquad & i, j = 1, \cdots, m, \\ \textstyle\sum_j p_{ij} = 1. \qquad & \end{aligned}$$

The probability structure of the chain is determined by P and the initial probability distribution vector $p = (p_i)$

$$(2) \qquad \begin{aligned} p_i = P[X(0) = i] \geqq 0, \qquad & i = 1, \cdots, m, \\ \textstyle\sum_i p_i = 1. \qquad & \end{aligned}$$

Suppose the experimenter does not observe the process $X(n)$ but rather a derived process $Y(n) = f(X(n))$ where f is a given function on $1, \cdots, m$. The states i of the original process $X(n)$ on which f equals some fixed constant are collapsed into a single state of the new process $Y(n)$. Call these collapsed sets of states $S_i, i = 1, \cdots, r, r \leqq m$. A natural question that arises is as to whether or not the new process is Markovian. It is clear that this is not generally the case.

Let us restrict ourselves to a process $X(n)$ with its initial probability distribution a left invariant vector of the matrix P, that is, $pP = p$. Further assume that all the components of p are positive (all transient states are thrown out). Let D be the diagonal matrix with its ith diagonal entry p_i. The process is said to be reversible if

$$DP = P'D$$

(P' is the transpose of P). The following result is obtained:

Theorem 1. *Let $X(n)$ be a stationary reversible process with $p_i > 0$ for all i. Then $Y(n)$ is Markovian if and only if for any fixed $\beta = 1, \cdots, r$*

$$(3) \qquad \sum_{j \varepsilon S_\beta} p_{ij} = P[X(n+1) \varepsilon S_\beta \mid X(n) = i] = C_{S_\alpha, S_\beta}$$

has the same value for all i in any given collapsed set of states S_α, $\alpha = 1, \cdots, r$.[2]

A slightly different problem can be phrased in the following way. Let

$$w = (w_i), w_i > 0, i = 1, \cdots, m$$

Received November 13, 1957; revised March 31, 1958.

[1] Work sponsored in part by the Office of Naval Research under Contract Nonr-908(10). Reproduction in whole or in part is permitted for the purpose of the United States Government.

[2] J. L. Snell pointed out that the original proof, given for Markov processes $X(n)$ with a symmetric P, holds for the reversible processes.

be any initial probability distribution. Consider the Markov process $X(n)$ generated by initial distribution w and transition probability matrix P. Again consider $Y(n) = f(X(n))$ and require that $Y(n)$ be Markovian whatever the initial distribution w.

COROLLARY 1. *A sufficient condition that $Y(n)$ be Markovian whatever the initial distribution w of $X(n)$ is given by* (3). Nonetheless, condition (3) is not generally necessary if the collapsed process is to be Markovian even in the problem covered in Corollary 1.

THEOREM 2. *Let f be a function that collapses only one class of states S. $Y(n)$ is Markovian whatever the initial distribution w of $X(n)$ if and only if one of the following two conditions is satisfied:*

(4) (i) $$\sum_{l \varepsilon S} p_{kl}\, p_{lu} = p_{k,S}\, C_u$$

for all $u \varepsilon S$ and all k;

(5) (ii) $$p_{i,S} = 0 \quad \text{for all} \quad i \varepsilon S.$$

Here

$$p_{k,S} = \sum_{j \varepsilon S} p_{kj} = P[X(n+1) \varepsilon S \mid X(n) = k].$$

An example of a Markov chain satisfying (4) but not (3) is given in the body of the paper.

Condition (4) naturally suggests the condition given in Corollary 2.

COROLLARY 2. *A sufficient condition that $Y(n)$ be Markovian, whatever the initial distribution w of $X(n)$, is given by*

(4') $$\sum_{l \varepsilon S_\alpha} p_{kl}\, p_{l,S_\beta} = p_{k,S_\alpha}\, C_{S_\alpha, S_\beta}$$

for all k, α, β.

Suppose we now go back and consider the class of stationary Markov chains $X(n)$ with $p_i > 0$, $i = 1, \cdots, m$, such that $Y(n) = f(X(n))$ is Markovian for any many-one transformation f.

THEOREM 3. *Let $X(n)$ be a stationary Markov chain with $p_i > 0$ $i = 1, \cdots, m$. $f(X(n))$ is Markovian for every many-one transformation f if and only if the transition probability matrix P of $X(n)$ is of the form*

(6) $$P = \alpha I + (1 - \alpha)U,$$

where U is a matrix with identical rows and α is a real number.

It is interesting to note that when one goes to the case of a decent continuous parameter Markov chain with a finite number of states, the analogue of (3) becomes almost necessary for $Y(t)$ to be Markovian, whatever the initial probability distribution w of $X(t)$.

THEOREM 4. *Let $X(t)$, $0 \leqq t < \infty$, be a Markov chain with a finite number of states $i = 1, \cdots, m$ and stationary transition probability function*

$$P(t) = (p_{ij}(t))$$

$$p_{ij}(t) = P[X(t + \tau) = j \mid X(\tau) = i]$$

continuous in t. Assume that

$$\lim_{t \downarrow 0} P(t) = I.$$

Clearly

$$P(t)P(s) = P(t + s), \qquad t, s > 0.$$

Let the initial probability distribution of $X(t)$ *be* w, $w_i > 0$, $i = 1, \cdots, m$. *Then* $Y(t) = f(X(t))$ *is Markovian, whatever the initial distribution* w *of* $X(t)$, *if and only if for each* $\beta = 1, \cdots, r$ *separately either*

$$(7) \qquad \begin{aligned} &\text{(i)} \quad p_{i,S_\beta}(t) \equiv 0 \quad \text{for all} \quad i \varepsilon S_\beta \quad \text{or} \\ &\text{(ii)} \quad p_{i,S_\gamma}(t) = C_{S_\beta,S_\gamma}(t) \quad \text{for every} \quad i \varepsilon S_\beta \quad \text{and all} \quad \gamma = 1, \cdots, r. \end{aligned}$$

Part of the interest in the proofs of Theorems 1 and 4 lies in the fact that they show that if the collapsed processes in these cases satisfy the Chapman-Kolmogorov equations, they are Markovian.

Condition (3) can be reworded in the case of a Markov process $X(t)$, $0 \leq t < \infty$, with stationary transition probabilities and values in an abstract space. Let Ω be a space of points x and $B(\Omega)$ a Borel field on Ω. Further let the sets (x) be elements of $B(\Omega)$. Consider a function

$$P(t; x, A), \qquad A \varepsilon B(\Omega)$$

satisfying

(i) $P(t; x, A)$ is a Baire function of x for fixed t, A;
(ii) $P(t; x, A)$ is a probability measure in $A \varepsilon B(\Omega)$ for fixed t, x;
(iii) $P(t; x, A)$ satisfies the Chapman-Kolmogorov equation

$$P(t + \tau; x, A) = \int_\Omega P(t; y, A)P(\tau; x, dy), \qquad t, \tau > 0.$$

Let $X(t)$ be a Markov chain with $P(t; x, A)$ as its transition probability function. Let f be a function from Ω onto another space of points Ω'. The function f induces a Borel field of sets $B(\Omega') = f(B(\Omega))$ on Ω'. This consists of sets of the form $fA = (y \varepsilon \Omega' \mid y = f(x), x \varepsilon A)$, $A \varepsilon B(\Omega)$. Now consider the inverse images of sets in $f(B(\Omega))$. The class of sets of this form we call $f^{-1}f(B(\Omega))$ and it is a subBorel field of $B(\Omega)$ consisting of sets of the form

$$\{z \varepsilon \Omega \mid z = f^{-1}f(x), x \varepsilon A\}, \qquad A \varepsilon B(\Omega).$$

The analogue of condition (3) is simply that

$$(8) \qquad P(t; x, A), \qquad A \varepsilon f^{-1}f(B(\Omega))$$

be a Baire function of x with respect to $f^{-1}f(B(\Omega))$ for fixed t, A.

COROLLARY 3. $Y(t) = f(X(t))$ *is a Markov process, whatever the initial probability distribution of* $X(t)$, *if condition* (8) *is satisfied*. Condition (8) is discussed

in a paper of B. Rankin [4] as a sufficient condition for a collapsed Markovian process to be Markovian.

2. The stationary case. Let the assumptions of Theorem 1 be satisfied. The matrix of n-step transition probabilities of the process $Y(n)$ is of the form

$$Q^{(n)} = AP^nB = (q_{\alpha\beta}^{(n)}) = (P[X(t+n) \varepsilon S_\beta \mid X(t) \varepsilon S_\alpha]), \tag{9}$$

where A, B are $r \times m$ and $m \times r$ matrices respectively. The elements of B are of the form

$$b_{ij} = \begin{cases} 1 & \text{if } \quad i \varepsilon S_j, \\ 0 & \text{otherwise;} \end{cases}$$

while

$$A = (B'DB)^{-1}B'D, \tag{10}$$

where D is the diagonal matrix introduced above. If the new process is Markovian, the Chapman-Kolmogonov equation must be satisfied by the $Q^{(n)}$, that is,

$$Q^{(n)} = AP^nB = [Q^{(1)}]^n = (APB)^n, \qquad n = 2, 3, \cdots. \tag{11}$$

This condition can be reworded in an equivalent form

$$AP^nBAPB = AP^{n+1}B, \qquad n = 1, 2, 3, \cdots. \tag{12}$$

Note that

$$BAPB = PB \tag{13}$$

implies that (12) is satisfied. Condition (13) is just condition (3) expressed in matrix form when the assumptions of Theorem 1 are satisfied. We first verify that (3) implies that $Y(n)$ is Markovian. (To facilitate printing we sometimes write $\alpha(i)$ in place of α_i.) Clearly

$$\begin{aligned} P[Y(0) \varepsilon S_{\alpha(0)}, \cdots, Y(n) \varepsilon S_{\alpha(n)}] &= \sum_{j=0}^{n} \sum_{i_j \varepsilon S_{\alpha(j)}} p_{i_0} p_{i_0 i_1} \cdots p_{i_{n-1} i_n} \\ &= \Big(\sum_{i \varepsilon S_{\alpha(0)}} p_i\Big) C_{S_{\alpha(0)}, S_{\alpha(1)}} \cdots C_{S_{\alpha(n-1)}, S_{\alpha(n)}} \end{aligned}$$

and it is easily seen that

$$C_{S_\alpha, S_\beta} = P[Y(n+1) \varepsilon S_\beta \mid Y(n) \varepsilon S_\alpha].$$

The sufficiency of condition (3) is thus verified. Note that the sufficiency argument given above holds for the case of any initial distribution w and without the condition of reversibility. We thus have Corollary 1.

Let us now consider the necessity of condition (3) when $X(n)$ is reversible. If $Y(n)$ is Markovian the Chapman-Kolmogorov equations are satisfied by the $Q^{(n)}$ and we must have

$$Q^{(2)} = [Q^{(1)}]^2$$

or

$$AP(I - BA)PB = 0.$$

But this implies that

$$B'DP(I - BA)PB = 0.$$

Because of reversibility, this can be written as

$$B'P'D(I - BA)PB = 0.$$

Now $D(I - BA)$ is positive definite so that

$$D(I - BA) = R'R$$

for some $m \times m$ matrix R. Thus

$$(RPB)'(RPB) = 0$$

and

$$RPB = 0.$$

But then

$$R'RPB = D(I - BA)PB = 0$$

and hence

$$(I - BA)PB = 0.$$

It is worth while noting that the problems we consider are related to issues of aggregation and consolidation in multisector models of mathematical economics (see [5]). There one has a stochastic matrix P and an invariant vector

$$p, pP = p.$$

One asks for the types of aggregation under which the aggregated invariant vector is an invariant vector of the aggregated matrix. The aggregated matrix $Q = APB$ where B is defined as before and $A = (B'D_vB)^{-1}B'D_v$. Here D_v is the diagonal matrix with its ith diagonal element v_i. The aggregation is determined by the sets of states S_i and the vector $v = (v_i)$. The aggregated vector is pB. The question is then for what aggregation schemes the relation

$$pBQ = pB(B'D_vB)^{-1}B'D_vPB = pB$$

is valid. Conditions (3) and (6) turn out to be crucial in some of the results obtained in [5].

3. Any initial distribution. Let the assumptions of Theorem 2 be satisfied. We first show that (4) is sufficient. It is enough to show that

$$\begin{aligned} P[X(n) = i, X(n+1) \in S, \cdots, X(n+h) \in S, X(n+h+1) = j] \\ = P[X(n) = i]P[X(n+1) \in S \mid X(n) = i] \\ \cdots P[X(n+h) \in S \mid X(n+h-1) \in S] \\ P[X(n+h+1) = j \mid X(n+h) \in S] \end{aligned}$$

for any $j \varepsilon S$ and any i, since then $Y(n)$ is clearly Markovian. Note that (4) implies that

$$\sum_{l \varepsilon S} p_{kl}\, p_{l,S} = p_{k,S}\, C_S \tag{14}$$

for all k. By making use of (4) and (14) the following relation is obtained

$$P[X(n+h+1) = j, X(n+h) \varepsilon S, \cdots, X(n+1) \varepsilon S \mid X(n) = i]$$

$$= \sum_{k=1}^{h} \sum_{i_k \varepsilon S} p_{i,i_1}\, p_{i_1,i_2} \cdots p_{i_{k-1},i_k}\, p_{i_k,j}$$

$$= p_{i,S}(C_S)^{h-1} C_j\,.$$

But

$$C_j = P[X(n+1) = j \mid X(n) \varepsilon S], \qquad j \not\varepsilon S,$$

and

$$C_S = P[X(n+1) \varepsilon S \mid X(n) \varepsilon S].$$

An Argument paralleling the one given above indicates that (4′) implies that $Y(n)$ is Markovian so that we have Corollary 2. $Y(n)$ is obviously Markovian if (5) is satisfied.

Now consider the necessity of (4). Since $Y(n)$ is Markovian whatever the initial distribution w of $X(n)$, the transition probabilities of $Y(n)$ satisfy the Chapman-Kolmogorov equation. It may be that $p_{iS} = 0$ for all i. Then (4) is obviously satisfied. Suppose now that there is an i such that $p_{iS} \neq 0$. The Chapman-Kolmogorov equation then tells us that

$$p_{i,S} \frac{\sum_{l \varepsilon S} \sum_{k} w_k\, p_{kl}\, p_{lu}}{\sum_{k} w_k\, p_{kS}} = \sum_{l \varepsilon S} p_{il}\, p_{lu}$$

for all i, $u \not\varepsilon S$. If k is such that $p_{k,S} \neq 0$ then

$$p_{i,S} \sum_{l \varepsilon S} p_{kl}\, p_{lu} = p_{k,S} \sum_{l \varepsilon S} p_{il}\, p_{lu} \tag{15}$$

as is seen by letting $w_k \to 1$ and $w_l \to 0$, $l \neq k$. And if $p_{k,S} = 0$ (15) is obviously satisfied. Thus (15) holds for all k and all $i \not\varepsilon S$. If there is an $i \not\varepsilon S$ such that $p_{iS} \neq 0$ (15) is satisfied for all k and i. But this implies relation (4). There is still the possibility that $p_{i,S} = 0$ for all $i \not\varepsilon S$, namely condition (5).

In the context of Theorem 2 condition (3) implies that condition (4) is satisfied. However, the converse is not true. Consider the transition probability matrix

$$P = \begin{bmatrix} \frac{1}{3} & \frac{1}{6} & \frac{1}{3} & \frac{1}{6} & 0 \\ \frac{1}{6} & \frac{1}{3} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \\ \frac{1}{4} & 0 & \frac{1}{4} & \frac{1}{6} & \frac{1}{3} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 & \frac{1}{4} \\ 0 & \frac{1}{4} & 0 & \frac{1}{2} & \frac{1}{4} \end{bmatrix}.$$

Collapse the states 1, 2, 3 into a set S and leave the states 4, 5 alone. Note that (3) is not satisfied. But (4) is satisfied since

$$\frac{\sum_{l \varepsilon S} p_{kl} \, p_{lu}}{p_{k,S}} = \tfrac{1}{6}$$

for all $u \varepsilon S$ and all k.

4. Any function f. The answer obtained to the question posed in Theorem 3 is the same as the answer obtained in a similar problem posed by Bush, Mosteller and others [1]. The structure of interest in Bush and Mosteller's problem is not Markovian. Note that in our case we ask that $f(X(n))$ have the same structure (a Markovian structure) as $X(n)$ for any f and a specific initial probability vector, a left invariant vector p of P. Bush and Mosteller ask that $f(X(n))$ have the same structure as $X(n)$ for any f and any initial probability vector w.

Let us now prove Theorem 3. The condition imposed on the process will not be used in full strength. Just consider a consolidation in which two states j, k are consolidated into a set S and all other states are left the same. Let i, l be any indices distinct from j, k. Since the consolidated process is Markovian, its transition probabilities satisfy the Chapman-Kolmogorov equation and hence

$$p_{il}^{(2)} = \sum_{u=1}^{m} p_{iu} \, p_{ul} = \sum_{u \notin S} p_{iu} \, p_{ul} + (p_{ij} + p_{ik}) \frac{p_j \, p_{jl} + p_k \, p_{kl}}{p_j + p_k}. \tag{16}$$

Equation (16) can be reduced to the following convenient form

$$(p_{ij}p_k - p_{ik}p_j)(p_{jl} - p_{kl}) = 0. \tag{17}$$

Further, (17) implies that

$$[(p_jp_{jj} + p_kp_{kj})p_k - (p_jp_{jk} + p_kp_{kk})p_j](p_{jl} - p_{kl}) = 0. \tag{18}$$

First consider the case in which for all i $p_{ij}p_k = p_{ik}p_j$ for all j, $k \neq i$. But then

$$p_{ij} = (1 - \lambda_i)p_j, \qquad i \neq j,$$

$$\lambda_i = \frac{p_{ii} - p_i}{1 - p_i},$$

so that P is of the form

$$P = \Lambda + (I - \Lambda)U,$$

where Λ is a diagonal matrix with diagonal elements λ_i and U is a matrix with identical rows $(p_1, \cdots, p_n)$. If

$$(p_jp_{jj} + p_kp_{kj})p_k = (p_jp_{jk} + p_kp_{kk})p_j \tag{19}$$

for some pair of indices j, k it follows that $\lambda_j = \lambda_k$. If (19) does not hold for the pair j, k, (18) implies that $p_{jl} = p_{kl}$ for all $l \neq j$, k. But then $\lambda_j = \lambda_k$. Thus it follows that in this case $\lambda_1 = \lambda_2 = \cdots = \lambda_n$.

Now on the contrary assume there is a row i for which $p_{ij}p_k = p_{ik}p_j$ does not hold for all $j, k \neq i$. Given any $j \neq i$ consider all k for which we can find a sequence $j_1, \cdots, j_\alpha$ such that

$$p_{ij}p_{j_1} = p_{ij_1}p_j, \qquad p_{ij_1}p_{j_2} = p_{ij_2}p_{j_1}, \cdots, p_{ij_\alpha}p_k = p_{ik}p_{j_\alpha}.$$

There is a maximal set of such indices k (including j of course). There are at least two such sets. The collection of all such maximal sets are disjoint. Given any j in one such maximal set and any j' in another we must have

$$p_{jl} = p_{j'l} \tag{20}$$

for all $l \neq j, j'$ and

$$p_{jj'} + p_{jj} - p_{j'j} - p_{j'j'} = 0. \tag{21}$$

For convenience let us assume $i = 1$. Keeping (20) and (21) in mind, it is clear that for any fixed $j \neq 1$ the p_{kj}'s must be equal for all $k \neq 1, j$. Call this common value u_j. Thus all rows except possibly for the first must be of the form

$$p_{kj} = \lambda\delta_{kj} + u_j.$$

There are now two possibilities. Either $p_{ij}p_k = p_{ik}p_j$ for all $i \neq 1$ and all

$$j, k \neq i$$

or this is not the case. If not we must have $p_{ij} = \lambda\delta_{ij} + u_j$ for all i. Since p is an invariant vector $u_j = (1 - \lambda)p_j$. On the other hand if $p_{ij}p_k - p_{ik}p_j = 0$ for all $i \neq 1$ and $j, k \neq i$ then $u_j = (1 - \lambda)p_j$. The elements of the first row are as yet unknown. But again making use of the fact that p is a stationary distribution we see that $p_{ij} = \lambda\delta_{1j} + (1 - \lambda)p_j$.

5. Finite state space and continuous time. The proof of the sufficiency of condition (7) in the case of Theorem 4 parallels the proof of Corollary 1.

We now show that (7) is necessary. A transition probability matrix-valued function $P(t)$ satisfying the regularity conditions posed in the assumptions in Theorem 4 is of the form (see [2])

$$P(t) = \exp(Gt),$$

where $G = (g_{ij})$ is such that

$$g_{ij} \geqq 0, \qquad i \neq j,$$

$$\sum_{\substack{j=1 \\ j \neq i}}^{m} g_{ij} = -g_{ii}.$$

Let $w = (w_i)$, $w_i > 0$ be the initial distribution of $X(t)$. A necessary condition that the collapsed process be Markovian for an initial vector can be written down conveniently in matrix notation. As before, let

$$Q_w^{(t)} = (B'D_wB)^{-1}B'D_wP(t)B$$

denote the t-step transition probability matrix (from time zero to time t) for the collapsed process $Y(t)$ when the initial probability distribution vector of the original process $X(t)$ is w. If the collapsed process $Y(t)$ is Markovian $Q_w^{(t)}$ must satisfy the Chapman-Kolmogorov equation and thus

$$Q_w^{(t)} Q_{wP(t)}^{(\tau)} = Q_w^{(t+\tau)}, \qquad t, \tau > 0, \tag{22}$$

for all w, $w_i > 0$. It is clear that the w_i's only have to satisfy $w_i > 0$ and that the condition $\sum w_i = 1$ needn't be imposed. On differentiating relationship (22) with respect to τ at $\tau = 0$ we obtain

$$Q_w^{(t)}(B'D_{wP(t)}B)^{-1}B'D_{wP(t)}GB = (B'D_wB)^{-1}B'D_wP(t)GB. \tag{23}$$

Let us now differentiate (23) with respect to t at $t = 0$. We then have

$$\begin{aligned} B'D_wGB(B'D_wB)^{-1}B'D_wGB - (B'D_wB)^{-1}B'D_{wG}BB'D_wGB + B'D_{wG}GB \\ = B'D_wG^2B. \end{aligned}$$

This can be written more conveniently as

$$B'[D_wG - G_{wG}][B(B'D_wB)^{-1}(B'D_w) - I]GB = 0. \tag{24}$$

Let

$$w_{S_\alpha} = \sum_{i \varepsilon S_\alpha} w_i,$$

$$g_{i,S_\alpha} = \sum_{j \varepsilon S_\alpha} g_{ij}.$$

Condition (24) can be written down elementwise as

$$\begin{aligned} \sum_{i \varepsilon S_\alpha} \sum_\gamma w_i\, g_{i,S_\alpha}\, w_{S_\gamma}^{-1} \sum_{i \varepsilon S_\gamma} w_i\, g_{i,S_\beta} - \sum_{i \varepsilon S_\alpha} \sum_k w_i\, g_{ik}\, g_{k,S_\beta} \\ - \sum_i w_i\, g_{i,S_\alpha}\, w_{S_\alpha}^{-1} \sum_{i \varepsilon S_\alpha} w_i\, g_{i,S_\beta} + \sum_i w_i \sum_{k \varepsilon S_\alpha} g_{ik}\, g_{k,S_\beta} = 0. \end{aligned} \tag{25}$$

If we set $w_i = u_i h$, $i \varepsilon S_\alpha$, in (25) and then let $h \downarrow 0$, the following relation is obtained since the first two terms drop out

$$- \sum_{i \notin S_\alpha} w_i\, g_{i,S_\alpha}\, u_{S_\alpha}^{-1} \sum_{i \varepsilon S_\alpha} u_i\, g_{i,S_\beta} + \sum_{i \notin S_\alpha} w_i \sum_{k \varepsilon S_\alpha} g_{i,k}\, g_{k,S_\beta} = 0.$$

But this is valid if and only if

$$g_{i,S_\alpha}\, u_{S_\alpha}^{-1} \sum_{i \varepsilon S_\alpha} u_i\, g_{i,S_\beta} = \sum_{k \varepsilon S_\alpha} g_{i,k}\, g_{k,S_\beta}$$

for all $i \notin S_\alpha$. Further, since this holds for all u_i,

$$g_{i,S_\alpha}\, g_{j,S_\beta} = \sum_{k \varepsilon S_\alpha} g_{i,k}\, g_{k,S_\beta} \tag{26}$$

for all $i \notin S_\alpha$ and all $j \varepsilon S_\alpha$. There are only two alternatives that arise. If

$$g_{i,S_\alpha} = 0$$

for all $i \varepsilon S_\alpha$ relationship (26) is obviously satisfied (we then say that S_α satisfies (i)). Otherwise $g_{i,s_\alpha} \neq 0$ for some $i \varepsilon S_\alpha$ in which case g_{j,s_β} for each β is a constant for all $j \varepsilon S_\alpha$, that is,

$$g_{j,s_\beta} = K_{s_\alpha,s_\beta} \tag{27}$$

for all $j \varepsilon S_\alpha$, $\beta = 1, \cdots, r$ (we then say that S_α satisfies (ii)). The matrix G is said to satisfy (7) if for each α separately S_α satisfies either (i) or (ii). Note that if G satisfies (7) the nth power of G, $G^n = (g_{ij}^{(n)})$, satisfies (7) in a consistent manner, that is, S_α satisfies (i) for G^n if and only if S_α satisfies (i) for G. Since

$$P(t) = \exp(Gt) = \sum_{k=0}^{\infty} G^k t^k / k!$$

$P(t)$ satisfies (7). It should be noted that our proof has shown that the condition that the Chapman-Kolmogorov equation be satisfied by the collapsed process is enough to imply that the new process be Markovian. P. Levy [3] has shown that this is generally not the case.

6. Abstract state space. Consider a Markov process $X(t)$ with initial probability distribution

$$P[X(0) \varepsilon A] = P(A), \qquad A \varepsilon B(\Omega)$$

and transition probability function

$$P(t; x, A)$$

satisfying the assumptions of Corollary 3. Then $Y(t) = f(X(t))$ is a Markovian process with initial distribution

$$P[Y(0) \varepsilon A'] = P[X(0) \varepsilon f^{-1}(A')] = Q(A')$$

$A' \varepsilon f(B(\Omega))$, and transition probability function

$$\begin{aligned} Q(t; y, A') &= P[Y(t + \tau) \varepsilon A' \mid Y(\tau) = y] \\ &= P[X(t + \tau) \varepsilon f^{-1}(A') \mid X(\tau) \varepsilon f^{-1}(y)] \\ &= P(t; x, f^{-1}(A')),\ y \varepsilon \Omega',\ A' \varepsilon f(B(\Omega)), \end{aligned}$$

where x is such that $y = f(x)$. This follows immediately from condition (8).

It is interesting to note that one can generate new Markovian processes from old ones by setting up f so that it is consistent with the symmetries of the transition probability mechanism of the old process. Consider $X(t)$ Brownian motion on the line. Here the transition probability density is

$$P(t; x, y) = (2\pi t)^{-1/2} \exp\left(-\frac{1}{2t}(x - y)^2\right), \qquad t > 0.$$

If we set

$$f(x) = x - a[x/a], \qquad a > 0,$$

where $[x]$ is the greatest integer less than or equal to x, the new Markovian process $Y(t) = f(X(t))$ is Brownian motion on the circle. If

$$f(x) = z$$

on all points of the form $2ka \pm z$, $0 \leqq z < a$, $k = 0, \pm 1, \cdots$, $Y(t)$ is Brownian motion on a line segment of length a with reflecting barriers at the endpoints.

As a further example consider starting out with two-dimensional Brownian motion $(X_1(t), X_2(t))$, that is, the transition probability density is

$$p(t; (x_1, x_2), (y_1, y_2)) = (2\pi t)^{-1} \exp\left(-\frac{1}{2t}\left[\left(x_1 - y_1\right)^2 + \left(x_2 - y_2\right)^2\right]\right), \quad t > 0.$$

If

$$f(x_1, x_2) = (u_1, u_2)$$

for all points (x_1, x_2) of the form $(u_1 + ja, u_2 + ka)$ $0 \leqq u_1, u_2 < a$, $j, k = 0, \pm 1, \cdots$ $(Y_1(t), Y_2(t))$ is Brownian motion on a torus. If

$$f(x_1, x_2) = (u_1, u_2)$$

for all points of the form $(u_1 + ja, (2k + j)\, a \pm u_2)$ $0 \leqq u_1, u_2 < a$, $j, k = 0, \pm 1, \cdots$ $(Y_1(t), Y_2(t))$ is Brownian motion on a Moebius strip with reflecting barriers on the edges of the strip.

REFERENCES

[1] BUSH AND MOSTELLER, *Stochastic Models for Learning*, John Wiley & Sons, New York.

[2] J. L. DOOB, *Stochastic Processes*, John Wiley & Sons, 1953.

[3] P. LEVY, "Examples de processus pseudo-Markoviens," *C. R. Acad. Sci. Paris*, Vol. 228 (1949), p. 2004–2006.

[4] B. RANKIN, "The concept of enchainment—a relation between stochastic processes" (1955), unpublished.

[5] D. ROSENBLATT, "On aggregation and consolidation in linear systems", to be published in the *Naval Research Logistics Quarterly*.

ASYMPTOTIC DISTRIBUTIONS OF "PSI-SQUARED" GOODNESS OF FIT CRITERIA FOR m-TH ORDER MARKOV CHAINS[1]

By Leo A. Goodman

University of Chicago

1. Introduction and Summary. Let $\{X_1, X_2, \cdots, X_N\}$ be an observed sequence from a stochastic process, where X_i can take any one of s values 1, 2, $\cdots$, s. Let f_{u} be the frequency of the m-tuple $\mathrm{u} = (u_1, u_2, \cdots, u_m)$ in the sequence. Let H_n' be the composite hypothesis that the process is a Markov chain of order n. Let H_n be any simple hypothesis belonging to H_n'. Let H_n^* be the maximum likelihood H_n. Let the expected value of f_{u} in a new sequence of length N given H_n be $f_{\mathrm{u},n}$, and given H_n^* be $f_{\mathrm{u},n}^*$. Let

$$\psi_{m,n}^2 = \sum_{\mathrm{u}} (f_{\mathrm{u}} - f_{\mathrm{u},n})^2/f_{\mathrm{u},n},$$

$$\psi_{m,n}^{*2} = \sum_{\mathrm{u}} (f_{\mathrm{u}} - f_{\mathrm{u},n}^*)^2/f_{\mathrm{u},n}^*,$$

$$\psi_{n+1,n}^{*2} = 0 .$$

Good had proposed in [7] the following two conjectures: (a) that the asymptotic distribution ($N \to \infty$) of $\psi_{m,n}^{*2}$, when H_n' is true, is

$$\mathop{*}_{\lambda=1}^{m-n-1} K_{g(\lambda)}(x/\lambda),$$

where $*$ denotes convolution, $g(\lambda) = (s-1)^2 s^{m-1-\lambda}$, and $K_i(x)$ is the χ^2-distribution with i degrees of freedom; (b) that the asymptotic distribution of $\psi_{m,n}^2$, when H_n is true, is

$$\mathop{*}_{\lambda=1}^{m-1} K_{g(\lambda)}(x/\lambda) * K_{s-1}(x/m),$$

mathematically independent of n. Conjectures (a) and (b) were proved by Billingsley [2] for the special case $n = 0$. For the special case $n = -1$ (by convention, H_{-1}' is the hypothesis of equiprobable or perfect randomness (see [7])), Conjecture (b) was proved by Good [5] when s is prime. In the present paper, Conjecture (a) will be proved for the general case $n \geqq -1$; conjecture (b) will be shown to be incorrect for $n > 0$, although a modified version of (b) will be proved for $n \geqq -1$. A third conjecture by Good [6] will also be proved here. It was

Received December 13, 1957.

[1] Research carried out at the Statistical Research Center, University of Chicago, under sponsorship of the Statistics Branch, Office of Naval Research. Reproduction in whole or in part is permitted for any purpose of the United States Government.

I am indebted to I. J. Good for the opportunity to read [7] before its publication, and also for mentioning that he suspected that his conjectures in [7] could be proved with the aid of the results in my earlier paper [10].

assumed in these earlier papers, and it will be assumed here, that all transition probabilities in the Markov chain are positive; the results can be modified accordingly when some of these probabilities are zero (see [1] and [10]).

Let $M_{m,n} = -2 \log \lambda_{n,m-1}$, where $\lambda_{n,m-1}$ is the ratio of the maximum likelihood given H'_n to that given H'_{m-1} (see [6]). For $m = n + 2$, the statistic $\psi^{*2}_{m,n}$ is asymptotically equivalent, when H'_n is true, to the likelihood ratio statistic $M_{m,n}$. For $m > n + 2$, $\psi^{*2}_{m,n}$ is asymptotically equivalent, when H'_n is true, to $\sum_{\lambda=1}^{m-n-1} \lambda M_{m+1-\lambda,m-1-\lambda}$, while $M_{m,n}$ is asymptotically equivalent to

$$\sum_{\lambda=1}^{m-n-1} M_{m+1-\lambda,m-1-\lambda}$$

(see [6], [10]). Thus, $\psi^{*2}_{m,n}$ corresponds asymptotically to a weighted sum of the likelihood ratio statistics $M_{n+2,n}, M_{n+3,n+1}, \cdots, M_{m,m-2}$, with the weights $m - n - 1, m - n - 2, \cdots, 1$, respectively, while $M_{m,n}$ weights these statistics equally (see [13] and reference to [13] in Section 4 herein).

Let $L_{m,n} = -2 \log \mu_{n,m-1}$, where $\mu_{n,m-1}$ is the ratio of the likelihood given H_n to the maximum likelihood given H'_{m-1}. For $m - 1 = n = 0$, the statistic $\psi^2_{m,n}$ is asymptotically equivalent, when H_n is true, to $L_{m,n}$. For $m - 1 > n = 0$, $\psi^2_{m,n}$ is asymptotically equivalent, when H_n is true, to

$$\sum_{\lambda=1}^{m-1} \lambda M_{m+1-\lambda,m-1-\lambda} + mL_{n+1,n},$$

while $L_{m,n}$ is asymptotically equivalent to $\sum_{\lambda=1}^{m-1} M_{m+1-\lambda,m-1-\lambda} + L_{n+1,n}$. For $n > 0$, the relation between $\psi^2_{m,n}$ and the likelihood ratio statistics $L_{m,n}$ and $M_{m,n}$ is not so straightforward. However, a modification $\psi'^2_{m,n}$ of $\psi^2_{m,n}$ (see Section 6 herein) is asymptotically equivalent, when H_n is true, to $L_{m,n}$ for $m = n + 1$, and to $\sum_{\lambda=1}^{m-n-1} \lambda M_{m+1-\lambda,m-1-\lambda} + (m - n)L_{n+1,n}$ for $m > n + 1$; while the likelihood ratio statistic $L_{m,n}$ is asymptotically equivalent to

$$\sum_{\lambda=1}^{m-n-1} M_{m+1-\lambda,m-1-\lambda} + L_{n+1,n}.$$

In [10], the m-tuple u was "split" into an $(m - n - 1)$-tuple, an n-tuple, and a 1-tuple; thus obtaining s^n "contingency tables" $(n \geq 0)$ each $s^{m-n-1} \times s$ (see [10]). The statistic $M_{m,n}$ can be seen to be asymptotically equivalent to the sum of the "likelihood ratio statistics" (for testing "independence" in each table) for the s^n tables, and the asymptotic distribution, when H'_n is true, of $M_{m,n}$ will be χ^2 with $s^n(s^{m-n-1} - 1)(s - 1) = s^m - s^{m-1} - s^{n+1} + s^n$ degrees of freedom. It is also possible to "split" the m-tuple u into an $(m - n - 1 - r)$-tuple, an n-tuple, and a $(1 + r)$-tuple $(0 \leq r \leq m - n - 2)$; thus obtaining s^n "contingency tables," each $s^{m-n-1-r} \times s^{1+r}$ (see [10]). The sum ${}_rM_{m,n}$ of the likelihood ratio (or any equivalent goodness of fit) statistics for the s^n tables will have an asymptotic mean value, when H'_n is true, of

$$s^n(s^{m-n-1-r} - 1)(s^{1+r} - 1) = s^m - s^{m-r-1} - s^{n+1+r} + s^n.$$

but the asymptotic distribution will not be χ^2 unless $r = 0$ or $m - n - 2$. It can be seen, using the methods developed in the present paper, that the statistic ${}_rM_{m,n}$ will be asymptotically equivalent, when H'_n is true, to

$$\sum_{\lambda=1}^{m-n-1} h(\lambda) M_{m+1-\lambda, m-1-\lambda},$$

where

$$h(\lambda) = \begin{cases} \lambda & \text{for} \quad 0 < \lambda \leqq v \\ v & \text{for} \quad v \leqq \lambda \leqq m - n - v \\ (m - n - \lambda) & \text{for} \quad m - n - v \leqq \lambda \leqq m - n - 1, \end{cases}$$

and $v = \min [r + 1, m - n - r - 1]$. Thus, the asymptotic distribution $(N \to \infty)$ of ${}_rM_{m,n}$ (or the corresponding asymptotically equivalent goodness of fit statistics), when H'_n is true, is

$$\mathop{*}_{\lambda=1}^{m-n-1} K_{g(\lambda)}[x/(h(\lambda))].$$

This result generalizes the earlier published results concerning the asymptotic distribution of the likelihood ratio statistic $M_{m,n}$ (or the corresponding asymptotically equivalent goodness of fit statistics) for testing the null hypothesis H'_n within H'_{m-1}, since ${}_rM_{m,n}$ for $r = 0$ or $m - n - 2$ is asymptotically equivalent to $M_{m,n}$ (see [6], [10]). A proof of this result will not be given since the method of proof is quite similar to that presented here for the asymptotic distribution of $\psi^{*2}_{m,n}$.

2. The Case $n = -1$**.** Let us first consider the case of equiprobable or perfect randomness ($n = -1$). We have that $H'_{-1} = H_{-1} = H^*_{-1}$, and $\psi^2_{m,-1} = \psi^{*2}_{m,-1}$. Thus, Conjectures (a) and (b) must be in agreement when $n = -1$. For $n = -1$, Conjecture (a) states that the asymptotic distribution of $\psi^{*2}_{m,-1}$ is

$$\mathop{*}_{\lambda=1}^{m} K_{g(\lambda)}(x/\lambda),$$

while (b) states that the asymptotic distribution of $\psi^2_{m,-1}$ is

$$\mathop{*}_{\lambda=1}^{m-1} K_{g(\lambda)}(x/\lambda) * K_{s-1}(x/m).$$

Thus, we must define $K_{g(m)}(x/m)$ as $K_{s-1}(x/m)$; i.e., $K_{(s-1)s^{s-1}}(x/m)$ as $K_{s-1}(x/m)$. It should also be mentioned that $\psi^2_{m,n}$ and $\psi^{*2}_{m,n}$ are defined only for $m \geqq n + 1$ (with $m \geqq 1$, for $n = -1$), and the symbol $\mathop{*}_{\lambda=1}^{0} K$ is to be understood as the atomic distribution that has the total probability 1 at the value $x = 0$. Since H'_{-1} is a special case of H'_0, results for $n = -1$ will follow directly from results for $n = 0$.

3. The Case $n = 0$. In the present paper, it will be convenient deal to with circular sequences, so that (for $m = 2$) $\sum_j f_{ij} = \sum_j f_{ji} = f_i$. A more general statement (for $m \geqq 2$) can be seen to hold for circular sequences (see [6]). A method of modifying results obtained for circular sequences so that they can be applied to linear sequences has been given in [9], and this method can be used to indicate that results analogous to those presented in the present paper will hold for linear sequences. The reader is cautioned that formulas for circular sequences can not be applied directly to linear sequences (see [9] and Corrigenda to [6]). It will also be convenient herein to replace $\psi^2_{m,n}$ and $\psi^{*2}_{m,n}$ by their asymptotically equivalent forms (when H_n is true)

$$\psi^2_{m,n} \approx 2\sum_u f_u \log (f_u / f_{u,n}),$$

and $\psi^{*2}_{m,n} \approx 2\sum_u f_u \log (f_u / f^*_{u,n})$, respectively.

Let us first consider Conjecture (a) when $n = 0$. For $m = 1$, this conjecture is obviously correct. For $m = 2$, this conjecture was first stated in [8] and was proved by Dawson and Good [4] and by Goodman [10]. The analogous result for the asymptotically equivalent form of $\psi^{*2}_{2,0}$ was proved by Hoel [11].

For $m = 3$, Conjecture (a) states that

$$\psi^{*2}_{3,0} = \sum_{ijk} (f_{ijk} - f_i f_j f_k/N^2)^2/(f_i f_j f_k/N^2)$$
$$\approx \chi^2_{s(s-1)^2} + 2\chi^2_{(s-1)^2},$$

where the symbols χ^2_i denote independent random variables each having a chi-square distribution with i degrees of freedom. (The $f_i f_j f_k/N^2$ used above is not the exact expected value, but is an asymptotic approximation; such asymptotic approximations for expected values will be used throughout.) We have that

$$\psi^{*2}_{3,0} \approx 2 \sum_{ijk} f_{ijk} \log [f_{ijk}/(f_i f_j f_k/N^2)]$$
$$= 2 \sum_{ijk} f_{ijk} \log [f_{ijk}/(f_i f_{jk}/N)] + 2 \sum_{jk} f_{jk} \log [f_{jk}/(f_j f_k/N)] .$$

The second term in the sum is asymptotically $\chi^2_{(s-1)^2}$, by the result for $m = 2$. The first term in the sum can be split into two parts, thus obtaining

$$2 \sum_{ijk} f_{ijk} \log [f_{ijk}/(f_{ij} f_{jk}/f_j)] + 2 \sum_{ij} f_{ij} \log [f_{ij} /(f_i f_j/N)] .$$

By the results in [10], for the test of H'_1 within H'_2, the asymptotic distribution of the first part is $\chi^2_{s(s-1)^2}$; the asymptotic distribution of the second part is $\chi^2_{(s-1)^2}$ (by the results for $m = 2$). The first part is asymptotically independent of the second part. This can be seen from the fact that their sum has the same asymptotic behavior, under H'_0, as the standard likelihood ratio statistic used in testing independence in an $s^2 \times s$ contingency table (see the test of H'_0 within H'_2 in [10]), and the two parts in the sum are obtained in the same manner as the partitioning of the likelihood ratio for the contingency table into two independent parts (see p. 439 in [3] and the articles referred to therein; rigorous

proofs of some of the published results concerning partitioning of contingency tables are given in [12][2]). The first part is obtained by separating the s^2 rows into s sets of s rows, thus obtaining s contingency tables, each $s \times s$, and using the combined likelihood ratio for the s tables to obtain asymptotically a $\chi^2_{s(s-1)^2}$ distribution (which leads to a test of H_1' within H_2' in [10]); the second part is obtained by combining the s rows in each set to obtain an $s \times s$ contingency table, and using the likelihood ratio for this table to obtain asymptotically a $\chi^2_{(s-1)^2}$ distribution (which leads to a test of H_0' within H_1' in [10]). Since the second part of the first term in $\psi^{*2}_{3,0}$ is equal to the second term in $\psi^{*2}_{3,0}$, their sum is asymptotically $2\chi^2_{(s-1)^2}$. Thus we have proved that $\psi^{*2}_{3,0} \approx \chi^2_{s(s-1)^2} + 2\chi^2_{(s-1)^2}$.

For $m = 4$, Conjecture (a) states that

$$\psi^{*2}_{4,0} = \sum_{ijkl} (f_{ijkl} - f_i f_j f_k f_l/N^3)^2/(f_i f_j f_k f_l/N^3)$$
$$\approx \chi^2_{s^2(s-1)^2} + 2\chi^2_{s(s-1)^2} + 3\chi^2_{(s-1)^2}.$$

We have that

$$\psi^{*2}_{4,0} \approx 2 \sum_{ijkl} f_{ijkl} \log [f_{ijkl}/(f_i f_j f_k f_l/N^3)]$$
$$= 2 \sum_{ijkl} f_{ijkl} \log [f_{ijkl}/(f_i f_{jkl}/N)] + 2 \sum_{jkl} f_{jkl} \log [f_{jkl}/(f_j f_k f_l/N^2)].$$

The second term in the sum is $\psi^{*2}_{3,0}$ and is asymptotically $\chi^2_{s(s-1)^2} + 2\chi^2_{(s-1)^2}$, by Conjecture (a) for $m = 3$. The first term can be split into two parts, thus obtaining

$$2 \sum_{ijkl} f_{ijkl} \log [f_{ijkl}/(f_{ijk} f_{jkl}/f_{jk})] + 2 \sum_{ijk} f_{ijk} \log [f_{ijk}/(f_i f_{jk}/N)].$$

By the results in [10] for the test of H_2' within H_3', the first part is asymptotically $\chi^2_{s^2(s-1)^2}$; the second part is asymptotically $\chi^2_{(s^2-1)(s-1)}$ (by the results for $m = 3$). The two parts are asymptotically independent. This follows from the fact that their sum has the same asymptotic behavior, under H_0', as the standard likelihood ratio statistic used in testing independence in an $s^3 \times s$ contingency table (see the test of H_0' within H_3' in [10]), and the two parts in the sum are obtained in the same manner as the partitioning of the likelihood ratio for the contingency table into two independent parts. The first part is obtained by separating the s^3 rows into s^2 sets of s rows, thus obtaining s^2 contingency tables, each $s \times s$, and using the combined likelihood ratio for the s^2 tables to obtain $\chi^2_{s^2(s-1)^2}$ (which leads to a test of H_2' within H_3' in [10]); the second part is obtained by combining the s rows in each set to obtain an $s^2 \times s$ contingency table, and using the likelihood ratio for this table we get $\chi^2_{(s^2-1)(s-1)}$ (which leads to a test of H_0' within H_2' in [10]). Since the second part of the first term in $\psi^{*2}_{4,0}$ can be written as $\chi^2_{(s^2-1)(s-1)} = \chi^2_{s(s-1)^2} + \chi^2_{(s-1)^2}$ (see the results for $m = 3$), and since the second term in $\psi^{*2}_{4,0}$ is $\psi^{*2}_{3,0} \approx \chi^2_{s(s-1)^2} + 2\chi^2_{(s-1)^2}$ (where the $\chi^2_{s(s-1)^2}$ and the $\chi^2_{(s-1)^2}$

[2] I am indebted to T. W. Anderson for bringing [12] to my attention.

expressions are identical with those appearing in the second part of the first term), their sum is asymptotically $2\chi^2_{s(s-1)^2} + 3\chi^2_{(s-1)^2}$. We have thus proved that

$$\psi^{*2}_{4,0} \approx \chi^2_{s^2(s-1)^2} + 2\chi^2_{s(s-1)^2} + 3\chi^2_{(s-1)^2}.$$

For $m = 5, 6, \cdots$, the same method of proof applies for Conjecture (a) when $n = 0$; it is easy to see that $\psi^{*2}_{m,0}$ is asymptotically equivalent, under H_0', to a weighted sum of asymptotically independent likelihood ratio statistics.

Let us now consider Conjecture (b) when $n = 0$. We have

$$\begin{aligned}\psi^2_{m,n} &\approx 2 \sum_u f_u \log (f_u/f_{u,n})\\ &= 2 \{\sum_u f_u \log (f_u/f^*_{u,n}) + \sum_u f_u \log (f^*_{u,n}/f_{u,n})\}\\ &\approx \psi^{*2}_{m,n} + 2 \sum_u f_u \log (f^*_{u,n}/f_{u,n}).\end{aligned}$$

For $m = 1$, $f^*_{u,0} = f_u$, and the second term is $2\sum_i f_i \log (f_i/Np_i)$, which is asymptotically χ^2_{s-1} by the standard statistical theory for goodness of fit tests. For $m = 2$, the second term is

$$\begin{aligned}2 \sum_u f_u \log (f^*_{u,0}/f_{u,0}) &= 2 \sum_{ij} f_{ij} \log [(f_i f_j/N)/Np_i p_j]\\ &= 4 \sum_i f_i \log (f_i/Np_i),\end{aligned}$$

which is asymptotically $2\chi^2_{(s-1)}$. The first term $\psi^{*2}_{2,0}$ is asymptotically independent of the second. This follows from the fact that the sum of $\psi^{*2}_{2,0}$ and $2\sum f_i \log (f_i/Np_i)$ is the likelihood ratio obtained in testing the null hypothesis H_0 that the transition probabilities for the Markov chain are $p_{ij} = p_j = p^0_j$ (specified) within the hypothesis H_1' (i.e., $2\sum_{ij} f_{ij} \log (f_{ij} / f_i p_j) \approx \chi^2_{(s-1)^2} + \chi^2_{(s-1)} = \chi^2_{s(s-1)}$ (see [1])), and the two terms in the sum are obtained by partitioning the likelihood ratio into two independent parts (the independence of the two parts follows directly from an examination of the asymptotic behavior of the f_{ij} (see, e.g., [9])). The first part is asymptotically $\chi^2_{(s-1)^2}$ and tests the null hypothesis H_0' that $p_{ij} = p_j$ (unspecified) within H_1'; the second part is asymptotically $\chi^2_{(s-1)}$ and tests the null hypothesis H_0 that $p_j = p^0_j$ (specified) within H_0'. Thus, $\psi^2_{2,0} \approx \chi^2_{(s-1)^2} + 2\chi^2_{(s-1)}$.

For $m = 3$ the second term is

$$\begin{aligned}2 \sum_u f_u \log(f^*_{u,0}/f_{u,0}) &= 2 \sum_{ijk} f_{ijk} \log [(f_i f_j f_k/N^2)/Np_i p_j p_k]\\ &= 6 \sum_i f_i \log (f_i/Np_i),\end{aligned}$$

which is asymptotically $3\chi^2_{(s-1)}$. The first term is independent of the second, by a similar argument to that presented for $m = 2$. Thus,

$$\psi^2_{3,0} \approx \chi^2_{s(s-1)^2} + 2\chi^2_{(s-1)^2} + 3\chi^2_{(s-1)}.$$

For $m = 4, 5, 6, \cdots$, the same method of proof applies for Conjecture (b) when $n = 0$.

We have thus given an altogether different method for proving the results obtained in [2] for $n = 0$; the results in [2] were based on the theory of finite-dimensional vector spaces. Since H'_{-1} is a special case of H'_0, the results given in the present section also prove that Conjectures (a) and (b), when properly interpreted, are true for $n = -1$, which generalizes the result proved in [5] for $n = -1$ and s prime. The different method presented in the present paper may further the understanding of the results in [2] and [5].

4. The Case $n = 1$. Let us now consider Conjecture (a) when $n = 1$. For $m = 2$, the conjecture is obviously true. For $m = 3$, we have that

$$\psi_{3,1}^{*2} \approx 2 \sum_{ijk} f_{ijk} \log [f_{ijk}/(f_{ij} f_{jk}/f_j)] \approx \chi^2_{s(s-1)^2},$$

by the results in [10] for the test of H'_1 within H'_2. For $m = 4$, we have that

$$\begin{aligned}\psi_{4,1}^{*2} &\approx 2 \sum_{ijkl} f_{ijkl} \log \{f_{ijkl}/[f_{ij} f_{jk} f_{kl}/(f_j f_k)]\} \\ &= 2 \sum_{ijkl} f_{ijkl} \log [f_{ijkl}/(f_{ijk} f_{jkl}/f_{jk})] + 2 \sum_{ijk} f_{ijk} \log [f_{ijk}/(f_{ij} f_{jk}/f_j)] \\ &\qquad + 2 \sum_{jkl} f_{jkl} \log [f_{jkl}/(f_{jk} f_{kl}/f_k)].\end{aligned}$$

By the results in [10] for the test of H'_2 within H'_3, the first term in the sum is asymptotically $\chi^2_{s^2(s-1)^2}$, and the second term is asymptotically $\chi^2_{s(s-1)^2}$ (see $m = 3$). The first term is asymptotically independent of the second. This follows from the fact that their sum can be regarded as the combined likelihood ratio used in testing independence in s contingency tables, each $s^2 \times s$ (see the test of H'_1 within H'_3 in [10]), and the two terms in the sum are obtained by partitioning the likelihood ratio for each of the s tables into two independent parts. For each of the s tables, the first part is obtained by separating the s^2 rows into s sets of s rows, thus obtaining s new tables, each $s \times s$, and using the combined likelihood ratio for the total of s^2 tables to obtain $\chi^2_{s^2(s-1)^2}$ (which is a test of H'_2 within H'_3 in [10]); the second part, for each of the original s tables, is obtained by combining the s rows in each new table to obtain an $s \times s$ table, and using the likelihood ratio for this table (there are s such tables) we get $\chi^2_{s(s-1)^2}$ (which is a test of H'_1 within H'_2 in [10]). The third term in the sum is asymptotically $\chi^2_{s(s-1)^2}$ (see $m = 3$), and it is equal to the second term in the sum. Thus we have $\psi_{4,1}^{*2} \approx \chi^2_{s^2(s-1)^2} + 2\chi^2_{s(s-1)^2}$.

For $m = 5, 6, \cdots$, the same method of proof applies for Conjecture (a) when $n = 1$; $\psi_{m,1}^{*2}$ is asymptotically equivalent to a weighted sum of asymptotically independent likelihood ratio statistics, under H'_1.

Let us now consider Conjecture (b) when $n = 1$. We have that

$$\psi_{m,1}^2 \approx \psi_{m,1}^{*2} + 2 \sum_u f_u \log (f_{u,1}^*/f_{u,1}).$$

For $m = 2$, $f^*_{u,1} = f_u$; thus, the first term $\psi^{*2}_{m,1} = 0$, and the second term is $2\sum_{ij} f_{ij} \log (f_{ij} / Np_i p_{ij})$, where the p_i are the stationary probabilities for the first order Markov chain with constant transition matrix $P = [p_{ij}]$. Conjecture (b) states that $\psi^2_{2,1} \approx \chi^2_{(s-1)^2} + 2\chi^2_{(s-1)}$. We could write

$$\psi^2_{2,1} \approx 2 \sum_{ij} f_{ij} \log [f_{ij}/(f_i f_j/N)] + 2 \sum_{ij} f_{ij} \log [(f_i f_j/N)/(Np_i p_{ij})].$$

The first is not asymptotically $\chi^2_{(s-1)^2}$, except when $n = 0$; and the second term is not asymptotically $2\chi^2_{(s-1)}$, except when $n = 0$. It is easy to see that Conjecture (b) will not hold true for $n = 1$, nor for $n > 1$.

Conjecture (b) will now be modified and this modified version will be proved true. This modification, for the special case $n = 1$, was first mentioned to the author by P. Billingsley in a private communication. In this communication, he mentioned that he had also obtained independently a proof of Conjecture (a), for the case $n = 1$, by very different methods from those used in the present paper, and that his results for Conjecture (a) and the modified Conjecture (b), when $n = 1$, could be extended to the case when $n > 1$, although the detailed asymptotic distributions were not given in the more general case [13].

Let $\psi'^2_{m,1} = \sum_u (f_u - f'_{u,1})^2/f'_{u,1}$, where $f'_{u,1}$ is the expected value of f_u in a new sequence of length N given H_1 and f_{u_1}; i.e., $f'_{u,1} = f_{u_1}\Pi_{i=1}^{m-1} p_{u_i u_{i+1}}$. Then

$$\psi'^2_{m,1} \approx \psi^{*2}_{m,1} + 2 \sum_u f_u \log (f^*_{u,1}/f'_{u,1}).$$

When $m = 2$, the first term $\psi^{*2}_{2,1}$ in the sum is zero and the second term is

$$2 \sum_{ij} f_{ij} \log (f_{ij}/f_i\, p_{ij}),$$

which is asymptotically $\chi^2_{s(s-1)}$ (see [1]). Thus, the asymptotic distribution of $\psi'^2_{2,1}$ is $\chi^2_{s(s-1)}$.

When $m = 3$, the first term $\psi^*_{3,1}$ is asymptotically $\chi^2_{s(s-1)^2}$, and the second term is

$$2 \sum_{ijk} f_{ijk} \log (f_{ij} f_{jk}/f_j f_i\, p_{ij}\, p_{jk}) = 4 \sum_{ij} f_{ij} \log (f_{ij}/f_i\, p_{ij}),$$

which is asymptotically $2\chi^2_{s(s-1)}$. The first term leads to a test of H'_1 within H'_2, and the second term leads to a test of H_1 within H'_1; it can be seen that the two terms are asymptotically independent under H_1. Thus, for $m = 3$, the asymptotic distribution of $\psi'^2_{m,1}$, when H_1 is true, is

$$\mathop{*}_{\lambda=1}^{m-2} K_{g(\lambda)}(x/\lambda) * K_{s(s-1)}[x/(m-1)].$$

This result can be proved for $m \geqq 3$ by the same method as given here for $m = 3$. Thus, a modified version of Conjecture (b) holds true for $n = 1$.

5. The Case $n = 2$. Let us now consider Conjecture (a) when $n = 2$. For $m = 3$, the conjecture is obviously true. For $m = 4$, we have

$$\psi^{*2}_{4,2} \approx \sum_{ijkl} f_{ijkl} \log [f_{ijkl}/(f_{ijk}\, f_{jkl}/f_{jk})] \approx \chi^2_{s^2(s-1)^2},$$

by the results in [10]. For $m = 5$, we have

$$\begin{aligned}\psi_{5,2}^{*2} &\approx 2 \sum_{ijklm} f_{ijklm} \log [f_{ijklm}/(f_{ijk} f_{jkl} f_{klm}/f_{jk} f_{kl})] \\ &= 2 \sum_{ijklm} f_{ijklm} \log [f_{ijklm}/(f_{ijkl} f_{jklm}/f_{jkl})] + 2 \sum_{ijkl} f_{ijkl} \log [f_{ijkl}/(f_{ijk} f_{jkl}/f_{jk})] \\ &\qquad + 2 \sum_{jklm} f_{jklm} \log [f_{jklm}/(f_{jkl} f_{klm}/f_{kl})].\end{aligned}$$

By the results in [10], the first term in the sum is asymptotically $\chi^2_{s^3(s-1)^2}$, and the second term is asymptotically $\chi^2_{s^2(s-1)^2}$ (see $m = 4$). The first term is asymptotically independent of the second; this follows by an argument similar to those appearing earlier here. The third term in the sum is asymptotically

$$\chi^2_{s^2(s-1)^2}$$

(see $m = 4$), and it is equal to the second term in the sum. Thus, we have $\psi_{5,2}^{*2} \approx \chi^2_{s^3(s-1)^2} + 2\chi^2_{s^2(s-1)^2}$.

For $m = 6, 7, \cdots$, the same method of proof applies for Conjecture (a) when $n = 2$. Conjecture (b) will not be true for $n = 2$, as it was not for $n = 1$. A modification of Conjecture (b) for $n = 2$ will now be given, which is similar to, although different from, Billingsley's modification of this conjecture for the special case $n = 1$.

Let $\psi_{m,2}'^2 = \sum_u (f_u - f_{u,2}')^2/f_{u,2}'$, where $f_{u,2}'$ is the expected value of f_u in a new sequence of length N given H_2 and $f_{u_1u_2}$; i.e., $f_{u,2}' = f_{u_1u_2} \prod_{i=1}^{m-2} p_{u_iu_{i+1}u_{i+2}}$ where $p_{u_1u_2u_3}$ is the second order transition probability that $X_t = u_3$, given that $X_{t-1} = u_2$ and $X_{t-2} = u_1$. Then $\psi_{m,2}'^2 \approx \psi_{m,2}^{*2} + 2 \sum_u f_u \log (f_{u,2}^*/f_{u,2}')$. When $m = 3$, the first term $\psi_{3,2}^{*2}$ in the sum is zero, and the second term is $2 \sum_{ijk} f_{ijk} \log (f_{ijk}/f_{ij} p_{ijk})$, which is asymptotically $\chi^2_{s^2(s-1)}$ (see [1]). Thus, the asymptotic distribution of $\psi_{3,2}'^2$ is $\chi^2_{s^2(s-1)}$.

When $m = 4$, the first term $\psi_{4,2}^{*2}$ is asymptotically $\chi^2_{s^2(s-1)^2}$, and the second term is

$$2 \sum_{ijkl} f_{ijkl} \log (f_{ijk} f_{jkl}/f_{jk} f_{ij} p_{ijk} p_{jkl}) = 4 \sum_{ijk} f_{ijk} \log (f_{ijk}/f_{ij} p_{ijk}),$$

which is asymptotically $2\chi^2_{s^2(s-1)}$. The first term leads to a test of H_2' within H_3', and the second term leads to a test of H_2 within H_2'; it can be seen that the two terms are asymptotically independent under H_2. Thus, for $m = 4$, the asymptotic distribution of $\psi_{m,2}'^2$, when H_2 is true, is

$$\mathop{*}_{\lambda=1}^{m-3} K_{g(\lambda)}(x/\lambda) * K_{s^2(s-1)}[x/(m-2)].$$

This result can be proved for $m \geqq 4$ by the same method as given here for $m = 4$. Thus, a modified version of Conjecture (b) holds true for $n = 2$.

6. The General Case. The method of proof used in the preceding sections for $n = -1, 0, 1, 2$ can also be applied when $n = 3, 4, \cdots$. In this way, Conjecture (a) can be proved in the general case $n \geqq -1$ and the following modification of

Conjecture (b) also holds in the general case. Let $\psi'^2_{m,n} = \sum_u (f_u - f'_{u,n})^2/f'_{u,n}$, where $f'_{u,n}$ is the expected value of f_u in a new sequence of length N given H_n and $f_{u_1u_2\cdots u_n}$ ($n \geqq 1$). Then, the asymptotic distribution of $\psi'^2_{m,n}$, when H_n is true, is

$$\mathop{*}_{\lambda=1}^{m-n-1} K_{g(\lambda)}(x/\lambda) * K_{s^n(s-1)}[x/(m-n)].$$

If we define $\psi'^2_{m,0}$ as $\psi^2_{m,0}$, then Conjecture (b) for $n = 0$, is identical with the modified version, and it also holds true. For $n = -1$, H_{-1} is a special case of H'_0, and the modified version of Conjecture (b) can be applied with n taken as zero. The reader will note that the asymptotic distribution of $\psi'^2_{m,n}$ is not mathematically independent of n; neither was the asymptotic distribution of $\psi^{*2}_{m,n}$. The result presented here for $\psi'^2_{m,n}$ generalizes Billingsley's result for $n = 1$.

A direct proof of these results could be given for the general case; this was not done here, since the proof proceeds along the same lines as the earlier discussion herein, and the results may be simpler to understand by considering first $n = 0$, $m = 1, 2, 3, 4, \cdots$; $n = 1, m = 2, 3, 4, \cdots$; $n = 2, m = 3, 4, \cdots$; etc.

In closing, we mention another conjecture by I. J. Good. In [6], the author conjectures that, when H'_{m-1} is true, the variables $-2 \log \lambda_{m-1,m}$ ($m = 0, 1, 2, \cdots$) are asymptotically independent, where $\lambda_{m-1,m}$ is the ratio of the maximum likelihood given H'_{m-1} to that given H'_m. If this conjecture were true, than an elegant proof of some results for testing H'_m within H'_n would be available (see [6]). We have that $-2 \log \lambda_{m-1,m} \approx \psi^{*2}_{m+1,m-1}$, when H'_{m-1} is true. The asymptotic independence of the likelihood ratios follows by the same kind of argument presented earlier in the present paper for the independence of some of the statistics considered (see, e.g., the reason why $\psi^{*2}_{4.2}$ and $\psi^{*2}_{3.1}$ are asymptotically independent, given $n = 1$, in the discussion here of the case $m = 4$ and $n = 1$).

The reader is referred to [13] for results that are closely related to some of those presented here, although the general approach and methods are very different. Also, some of the work in [14], [15], and [16] has some (but not much) relation to the present paper.

REFERENCES

[1] M. S. Bartlett, "The frequency goodness of fit test for probability chains," *Biometrika*, Vol. 47 (1951), pp. 86–95.

[2] P. Billingsley, "Asymptotic distributions of two goodness of fit criteria," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 1123–9.

[3] W. G. Cochran, "Some methods for strengthening the common χ^2 tests," *Biometrics*, Vol. 10 (1954), pp. 417–51.

[4] R. Dawson and I. J. Good, "Exact Markov probabilities from oriented linear graphs," *Ann. Math. Stat.*, Vol. 28 (1957), pp. 838–946.

[5] I. J. Good, "The serial test for sampling numbers and other tests for randomness," *Proc. Cambridge Philos. Soc.*, Vol. 49 (1953), pp. 276–84.

[6] I. J. Good, "The likelihood ratio test for Markoff chains," *Biometrika*, Vol. 42 (1955), pp. 531–3; Corrigenda, Vol. 44 (1957), p. 301.

[7] I. J. Good, "Review of P. Billingsley's 'Asymptotic distributions of two goodness of fit criteria'," *Mathematical Reviews*, Vol. 18 (1957), p. 607.

[8] LEO A. GOODMAN, "On the statistical analysis of Markov chains" (abstract), *Ann. Math. Stat.*, Vol. 26 (1955), p. 771.

[9] LEO A. GOODMAN, "Simplified runs tests and likelihood ratio tests for Markoff chains," *Biometrika*, Vol. 45 (1958), pp. 181–97.

[10] LEO A. GOODMAN, "Exact probabilities and asymptotic relationships for some statistics from m-th order Markov chains," *Ann. Math. Stat.*, Vol. 29 (1958), pp. 476–90.

[11] P. G. HOEL, "A test for Markoff chains," *Biometrika*, Vol. 41 (1954), pp. 430–3.

[12] MARILYN SPENCER, "The two-way contingency table and approximate chi-square distributions," Master's Essay, Columbia University Library, 1957.

[13] P. BILLINGSLEY, "On testing Markov Chains," unpublished report.

[14] G. P. BASHARIN, "On using the test of goodness of fit (χ^2 agreement) as a test of independence of trials," *Teoriya Veroyatœnostei i ee Primereniya*, Vol. 2 (1957), pp. 141–2.

[15] N. V. SMIRNOV, "The statistical estimation of transition probabilities in Markov chains," *Vestnik Leningradskoyo Universiteta*, Vol. 1 (1955), pp. 47–8.

[16] V. E. STEPANOW, "Some statistical tests for Markov chains," *Teoriya Veroyatnostei i ee Primeneniya*, Vol. 2 (1957), pp. 143–4.

EFFICIENCY PROBLEMS IN POLYNOMIAL ESTIMATION

By Paul G. Hoel

University of California, Los Angeles

1. Summary. Using the generalized variance as a criterion for the efficiency of estimation, the best choice of fixed variable values within an interval for estimating the coefficients of a polynomial regression curve of given degree is determined for the classical regression model. Using this same criterion, some results are obtained on the increased efficiency arising from doubling the number of equally spaced observation points

(i) when the total interval is fixed and

(ii) when the total interval is doubled. Measures of the increased efficiency are found for the classical regression model and for models based on a particular stationary stochastic process and a pure birth stochastic process.

2. Introduction. In the classical theory of regression, a set of values $x_1, x_2, \cdots, x_n$ of a variable x is selected and observations are made on a related variable y corresponding to those selected x values. If y_i denotes the y value corresponding to x_i, it is then assumed that $y_1, y_2, \cdots, y_n$ are uncorrelated variables with a common variance σ^2. Now if it is assumed that the means of the y's lie on a polynomial curve of degree k, that is, that

$$E(y_i) = \beta_0 + \beta_1 x_i + \cdots + \beta_k x_i^k \tag{1}$$

then a basic problem in statistics is how best to estimate the β's.

There are two aspects to this estimation problem. One is to determine the best method for using the information given by a set of n observtions $y_1, y_2, \cdots, y_n$. The other is to determine the best method for choosing the x values at which to take observations.

Although much research has gone into studying the first aspect of the problem, considerably less has been done on the second. Many years ago, K. Smith [1] was able to determine those x values within a fixed interval that minimize the maximum variance of a single estimated ordinate for polynomials up to degree six. More recently, De La Garza [2] was able to show that just as much information is obtained from observations made at certain $k + 1$ points in the interior of an interval as from n distinct points in that interval. Elfving [3], Chernoff [4], Daniels [5], and Ehrenfeld [6] have also made contributions toward this and other closely related problems.

In this paper an optimum solution based on the generalized variance is given for the problem of how to choose the x values in an interval for the classical regression model. In addition, a beginning is made on the more general problem

Received January 24, 1958; revised June 18, 1958.

of how to choose x values for efficient polynomial estimation when one drops the assumption that the y's are uncorrelated.

3. Estimation methods. When a number of parameters are to be estimated simultaneously, the volume of the ellipsoid of concentration of the estimates is often used as a measure of the efficiency of the estimates. Since the square of the volume of the ellipsoid of concentration is proportional to the generalized variance of the estimates, one can just as well use the generalized variance as a measure of efficiency. This is the measure that will be used in this paper for making comparisons of different sets of estimates.

Suppose one wished to estimate the function $\lambda_0\beta_0 + \lambda_1\beta_1 + \cdots + \lambda_k\beta_k$, where the λ's are an arbitrary set of real numbers, by means of a linear estimate, $c_1y_1 + c_2y_2 + \cdots + c_ny_n$. Suppose further that the estimate is to be unbiased and possess minimum variance. Then it can be shown that the resulting estimates for the β's are given by the matrix formula

$$\hat{\beta} = (X'S^{-1}X)^{-1}X'S^{-1}y \tag{2}$$

where S is the covariance matrix of the y's and X is the matrix

$$X = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^k \\ 1 & x_2 & x_2^2 & \cdots & x_2^k \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^k \end{bmatrix}.$$

Furthermore, it can also be shown that the generalized variance of these estimates is given by the determinant formula

$$\text{G.V.} = |X'S^{-1}X|^{-1} \tag{3}$$

These same formulas will be obtained if one assumes that the y's possess a multivariate normal distribution and then finds the maximum likelihood estimates of the β's.

The advantage of the estimates given by formula (2) lies in the fact that it can be shown that among all linear unbiased estimates of the β's, the estimates given by this formula possess a minimum generalized variance. Thus, if one restricts himself to linear estimates, these are optimum estimates. All the comparisons to be made in the following sections will assume that the estimates are those given by formula (2), and hence that the generalized variance is given by formula (3).

4. Classical regression. Since the classical regression model assumes that $y_1, y_2, \cdots, y_n$ are uncorrelated with a common variance σ^2, the covariance matrix S is a diagonal matrix with elements σ^2.

Now De La Garza [2] has shown that the same information matrix, $X'S^{-1}X$, and hence the same value of the generalized variance, can be obtained by replacing a given set of n observations at the points $x_1, x_2, \cdots, x_n$ by a total of n

observations made at $k + 1$ properly selected points in the interval from x_1 to x_n. These points will be denoted by $t_1, t_2, \cdots, t_{k+1}$ and the number of observations to be made at t_i will be denoted by n_i, where $\sum_{i=1}^{k+1} n_i = n$. In terms of these substitute observations, the matrices in (3) are all square matrices and therefore the determinant of their product can be obtained by taking the product of their determinants. As a result, (3) will assume the form

$$\frac{1}{\text{G.V.}} = \begin{vmatrix} 1 & 1 & \cdots & 1 \\ t_1 & t_2 & \cdots & t_{k+1} \\ \vdots & \vdots & & \vdots \\ t_1^k & t_2^k & \cdots & t_{k+1}^k \end{vmatrix} \begin{vmatrix} \frac{n_1}{\sigma^2} & 0 & \cdots & 0 \\ 0 & \frac{n_2}{\sigma^2} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \frac{n_{k+1}}{\sigma^2} \end{vmatrix} \begin{vmatrix} 1 & t_1 & \cdots & t_1^k \\ 1 & t_2 & \cdots & t_2^k \\ \vdots & \vdots & & \vdots \\ 1 & t_{k+1} & \cdots & t_{k+1}^k \end{vmatrix}$$

$$= \begin{vmatrix} 1 & 1 & \cdots & 1 \\ t_1 & t_2 & \cdots & t_{k+1} \\ \vdots & \vdots & & \vdots \\ t_1^k & t_2^k & \cdots & t_{k+1}^k \end{vmatrix}^2 \frac{\prod_{i=1}^{k+1} n_i}{\sigma^{2k+2}}.$$

But this determinant is a Vandermonde determinant with value $\prod_{i<j}(t_i - t_j)$; consequently

$$\frac{1}{\text{G.V.}} = \frac{1}{\sigma^{2k+2}} \prod_{i<j}^{k+1} (t_i - t_j)^2 \prod_{i=1}^{k+1} n_i \tag{4}$$

Since $\prod_{i=1}^{k+1} n_i$, subject to the restriction $\sum_{i=1}^{k+1} n_i = n$, is maximized when $n_1 = n_2 = \cdots = n_{k+1}$, it follows that the generalized variance will be minimized for a fixed set of values when the same number of observations is taken at each of the t values. This assumes that n will be chosen to make $n/(k + 1)$ an integer.

Now consider the maximization of $\prod_{i<j}(t_j - t_j)^2$, subject to the restriction that $x_1 \leqq t_i \leqq x_n$, $i = 1, \cdots, k + 1$. If x is transformed linearly so that this restriction assumes the form $-1 \leqq t_i \leqq 1$, $i = 1, \cdots, k + 1$, then it is known [7] that the set of t values that maximizes $\prod_{i<j}(t_i - t_j)^2$ is given by the zeros of a polynomial which is the integral of one of the Legendre polynomials. These zeros can be obtained from the proper tables [8].

It is clear from inspecting the function $\prod(t_i - t_j)^2$ that the end points of the interval will always be chosen as two of the t values. It is also clear that the greater the range of x values, the smaller will be the generalized variance.

In view of the preceding results, it follows that optimum linear estimates of the coefficients of classical polynomial regression are obtained by using the estimates given by formula (2), choosing as large a range of x values as possible, taking observations at the $k + 1$ points in this range given by means of the zeros of a tabulated polynomial, and repeating the experiment as many times as the total set, n, of observations will permit, with n chosen to make $n/(k + 1)$ an integer.

The preceding optimum manner of choosing x values assumes that the generalized variance of the estimates of the coefficients of the regression polynomial is the proper measure of efficiency to use. If the sample regression polynomial curve is to be used exclusively for estimating ordinates of the theoretical regression polynomial curve, then one might prefer a measure of efficiency based on the variances and covariances of such estimated values. From this point of view, let $\tau_1, \cdots, \tau_{k+1}$ denote $k + 1$ arbitrary points chosen in the given interval. Further, let α_i and $\hat{\alpha}_i$ denote the ordinate, and its estimate, of the polynomial regression curve at τ_i. Thus,

$$\alpha_i = \beta_0 + \beta_1\tau_i + \cdots + \beta_k\tau_i^k, \qquad i = 1, \cdots, k + 1$$

and

$$\hat{\alpha}_i = \hat{\beta}_0 + \hat{\beta}_1\tau_i + \cdots + \hat{\beta}_k\tau_i^k, \qquad i = 1, \cdots, k + 1.$$

Calculations will yield the covariance formula

$$m_{ij} = E(\hat{\alpha}_i - \alpha_i)(\hat{\alpha}_j - \alpha_j) = \sum_{r=0}^{k}\sum_{s=0}^{k} \sigma_{rs}\tau_i^r\tau_j^s$$

where σ_{rs} is the covariance of $\hat{\beta}_r$ and $\hat{\beta}_s$. Since the generalized variance is the determinant of the covariance matrix, the generalized variance of the $\hat{\alpha}$'s will be equal to the determinant $|m_{ij}|$. But it will be observed that the matrix (m_{ij}) can be written in the form

$$(m_{ij}) = \begin{bmatrix} 1 & \tau_1 & \cdots & \tau_1^k \\ 1 & \tau_2 & \cdots & \tau_2^k \\ \vdots & \vdots & & \\ 1 & \tau_{k+1} & \cdots & \tau_{k+1}^k \end{bmatrix} \begin{bmatrix} \sigma_{00} & \cdots & \sigma_{0k} \\ \vdots & & \vdots \\ \sigma_{k0} & \cdots & \sigma_{kk} \end{bmatrix} \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \tau_1 & \tau_2 & \cdots & \tau_{k+1} \\ \vdots & \vdots & & \vdots \\ \tau_1^k & \tau_2^k & \cdots & \tau_{k+1}^k \end{bmatrix}$$

Since $|\sigma_{rs}|$ is the generalized variance of the β's, it follows that

$$\text{G.V. } (\hat{\alpha}) = \text{G.V. } (\hat{\beta}) \prod_{i<j}^{k+1} (\tau_i - \tau_j)^2$$

This result shows that the generalized variance of the estimates of the ordinates of a polynomial regression curve at $k + 1$ arbitrary points in an interval will be minimized when the generalized variance of the estimates of the coefficients of the polynomial regression curve is minimized[1].

A recent paper by Guest [11], which was published after this paper had been submitted, has generalized the results of Smith [1] to polynomials of any degree. He shows that the values of $t_1, t_2, \cdots, t_n$ that minimize the maximum variance of a single estimated ordinate are given by means of the zeros of the derivative of a Legendre polynomial. It is easily seen that this set of values is the same set which minimizes the generalized variance above. Thus, whether one is interested

[1] I am indebted to Professor John Tukey for suggesting this relationship.

in efficient estimation of regression coefficients, or in efficient ordinate estimation, either at $k + 1$ points or one point, the optimum choice of t values is the same.

5. Comparison methods. When the assumption that the y's are uncorrelated is dropped, the problem of how best to choose the x's becomes very difficult. The choice will depend in a complicated manner upon the covariance matrix S. As a consequence, comparisons will be made only for equally spaced sets of points and only for three classes of covariance matrices. The sets of points that were selected for consideration are the following:

(1) n equally spaced points in the interval $(0, l)$

(2) $2n$ equally spaced points in the interval $(0, l)$

(3) $2n$ equally spaced points in the interval $(0, 2l)$

(4) two sets of observations of type (1).

A comparison of the relative advantages of choices (2), (3), and (4) over (1) will be made by comparing their generalized variances. Letting δ denote the interval between consecutive x values, these generalized variances will be denoted by G.V. (n, δ), G.V. $(2n, \delta/2)$, G.V. $(2n, \delta)$, and G.V. (2 runs), respectively.

The three classes of covariance matrices that will be studied are the following:

(a) uncorrelated variables, common variance

(b) $\rho(y_i, y_j) = e^{-\alpha|x_i - x_j|}$, $\alpha > 0$, common variance

(c) covariance matrix of a pure birth stochastic process.

The first of these is the classical regression model considered in the preceding section. The second is the covariance matrix of a particular stationary stochastic process. The third was selected because it represents a stochastic process of the non-stationary type and in which the covariances grow as x increases. These three covariance matrices cover a rather wide range of correlation relationships and therefore conclusions obtained from them should have a rather wide range of application.

For comparison purposes it is advantageous to consider the following three ratios:

$$
\begin{aligned}
R_1 &= \left[\frac{\text{G.V. } (n, \delta)}{\text{G.V. } (2n, \delta/2)}\right]^{1/(k+1)} \\
R_2 &= \left[\frac{\text{G.V. } (n, \delta)}{\text{G.V. } (2n, \delta)}\right]^{1/(k+1)} \qquad (5) \\
R_3 &= \left[\frac{\text{G.V. } (n, \delta)}{\text{G.V. } (m) \text{ runs}}\right]^{1/(k+1)}
\end{aligned}
$$

The reason for these choices is that it is easily shown that R_3 has the value m; consequently if the value of R_1, for example, should turn out to be m, it can be concluded that m runs of the basic experiment are needed to yield the same efficiency of estimation as that obtained by doubling the number of equally spaced observation points in the given interval. All comparisons will be made

in this manner, that is, by stating the number of runs of the experiment needed to yield the same efficiency as the choice of x values being considered.

6. Uncorrelated variables. It will be assumed that $n > k + 1$; consequently the X matrix in (3) will not be a square matrix and formula (4) will not be applicable. Under equal spacing in the interval $(0, l)$, the x values will be chosen as $x_i = i\delta$. As a result, the X matrix will assume the form

$$X = \begin{bmatrix} 1 & \delta & \cdots & \delta^k \\ 1 & 2\delta & \cdots & (2\delta)^k \\ \vdots & \vdots & & \vdots \\ 1 & n\delta & \cdots & (n\delta)^k \end{bmatrix} \tag{6}$$

Since S^{-1} is a diagonal matrix with elements $1/\sigma^2$, it is easily seen that (3) reduces to

$$\frac{1}{\text{G.V.}\ (n, \delta)} = \frac{\delta^{k(k+1)}}{\sigma^{2k+2}} \begin{vmatrix} n & \sum_1^n i & \cdots & \sum_1^n i^k \\ \sum_1^n i & \sum_1^n i^2 & \cdots & \sum_1^n i^{k+1} \\ \vdots & \vdots & & \vdots \\ \sum_1^n i^k & \sum_1^n i^{k+1} & \cdots & \sum_1^n i^{2k} \end{vmatrix}. \tag{7}$$

The value of this determinant is known [10] to be the polynomial displayed in (8); hence

$$\frac{1}{\text{G.V.}\ (n, \delta)} = \frac{\delta^{k(k+1)}}{\sigma^{2k+2}} An^{k+1}(n^2 - 1^2)^k(n^2 - 2^2)^{k-1} \cdots (n^2 - k^2) \tag{8}$$

where $A = (1!\,2! \cdots k!)^4 / (1!\,2! \cdots (2k+1)!)$. The value of R_1 given in (5) then becomes

$$R_1 = \frac{1}{2^k}\left[\frac{(2n)^{k+1}(4n^2 - 1^2)^k \cdots (4n^2 - k^2)}{n^{k+1}(n^2 - 1^2)^k \cdots (n^2 - k^2)}\right]^{1/k+1}. \tag{9}$$

Using (8) and (5), it follows readily that

$$R_2 = 2^k R_1. \tag{10}$$

Now consider the limiting values of R_1 and R_2 as $n \to \infty$. The resulting values may be considered as asymptotic measures of efficiency. From (9) and (10) it follows that

$$\lim_{n\to\infty} R_1 = 2 \quad \text{and} \quad \lim_{n\to\infty} R_2 = 2^{k+1}.$$

The first result implies that if one has a large number of equally spaced points in a fixed interval at which observations are made, then two runs of the experi-

ment will yield the same efficiency of estimation as doubling the number of equally spaced points in that interval. The second result implies, for example, that if the polynomial regression curve is of degree 4, then 32 runs of the experiment will be needed to yield the same efficiency of estimation as doubling the number of points by doubling the interval over which observations are to be made. It is clear from this second result that the higher the degree of the polynomial the more important it is to extend the range of x values as far as possible.

7. Stationary process model. Denoting the correlation between y_i and y_j by ρ_{ij}, it follows under equal spacing that the correlation function for model (b) will assume the form

$$\rho_{ij} = e^{-\alpha|x_i - x_j|} = e^{-\alpha\delta|i-j|}$$

Letting $w = e^{-\alpha\delta}$ and setting $\sigma^2 = 1$, since it will always cancel out in the R ratios, it will be seen that the covariance matrix here is given by

$$S = \begin{bmatrix} 1 & w & w^2 & \cdots & w^{n-1} \\ w & 1 & w & \cdots & w^{n-2} \\ \vdots & \vdots & \vdots & & \vdots \\ w^{n-1} & w^{n-2} & w^{n-3} & \cdots & 1 \end{bmatrix}$$

Calculations will show that the inverse of S is given by

$$S^{-1} = \frac{1}{1-w^2} \begin{bmatrix} 1 & -w & 0 & \cdots & 0 & 0 \\ -w & 1+w^2 & -w & \cdots & 0 & 0 \\ 0 & -w & 1+w^2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -w & 1 \end{bmatrix}$$

If S^{-1} is written as the sum of several matrices and then premultiplied by X' and postmultiplied by X, and finally brought together again into one matrix, it will be found that (3) assumes the form

$$\frac{1}{\text{G.V.}\ (n, \delta)} = \frac{\delta^{k(k+1)}}{(1-w^2)^{k+1}} \mid B(n, w) \mid \tag{11}$$

where $B(n, w)$ is the matrix whose element in row $p + 1$ and column $q + 1$ is given by

$$\begin{aligned} b_{p+1\ q+1} &= (w^2 + 1) \sum_1^n i^{p+q} \\ &\quad - w \sum_2^n [i^p(i-1)^q + i^q(i-1)^p] - w^2[n^{p+q} + 1]. \end{aligned} \tag{12}$$

Since $w = e^{-a\delta}$, the value of G.V. $(2n, \delta/2)$ can be obtained by replacing n by $2n$, δ by $\delta/2$, and w by $\sqrt{w}$ in (11). As a result, it will follow that

$$R_1 = \frac{1+w}{2^k} \frac{| B(2n, \sqrt{w}) |^{1/(k+1)}}{| B(n, w) |^{1/(k+1)}}.$$

Similarly,

$$R_2 = \frac{| B(2n, w) |^{1/(k+1)}}{| B(n, w) |^{1/(k+1)}}.$$

Now allow $n \to \infty$. From (12) it will be observed that the dominating part of $b_{p+1\,q+1}$ is $(w-1)^2 \sum i^{q+q}$. As a result, the asymptotic value of the determinant $| B(n, w) |$ is

$$\begin{vmatrix} (w-1)^2 n & (w-1)^2 \sum i & \cdots & (w-1)^2 \sum i^k \\ (w-1)^2 \sum i & (w-1)^2 \sum i^2 & \cdots & (w-1)^2 \sum i^{k+1} \\ \vdots & \vdots & & \vdots \\ (w-1)^2 \sum i^k & (w-1)^2 \sum i^{k+1} & \cdots & (w-1)^2 \sum i^{2k} \end{vmatrix}$$

But this is merely $(w-1)^{2k+2}$ times the determinant in (7), which in turn has the asymptotic value $An^{(k+1)^2}$. From the preceding results, it follows that

$$\lim_{n\to\infty} R_1 = \frac{w+1}{2^k} \frac{(\sqrt{w}-1)^2 2^{k+1}}{(w-1)^2} = \frac{2(w+1)}{(\sqrt{w}+1)^2}$$

and

$$\lim_{n\to\infty} R_2 = 2^{k+1}$$

For the purpose of seeing the implications of these formulas, consider the numerical value $w = e^{-a\delta} = .64$. This value implies that the correlation coefficient between neighboring y values is .64. Calculations yield the values

$$\lim_{n\to\infty} R_1 = 1.01 \quad \text{and} \quad \lim_{n\to\infty} R_2 = 2^{k+1}.$$

Thus, doubling the number of observation points in a given interval, when there are already a large number of such points, gives practically no additional estimation information. The value of R_2, however, shows that the same asymptotic efficiency is gained here as in the case of uncorrelated variables. For correlated variables like those being considered in this section, it is clear that the interval over which observations are to be made should be extended as far as possible, but that if it can't be extended, repeating the experiment is far more efficient than taking additional observation points.

8. Pure birth process model. Although a pure birth process is a discrete process with an exponential regression curve, it was selected only for its covariance matrix properties which are quite different from those of the two preceding models.

If b denotes the constant asymptotic birth rate, y_0 the population size at time t_0, and y the population size at time $t > t_0$, then the conditional probability function for y, given y_0, is

$$P\{y_0, y; t_0, t\} = \binom{y-1}{y_0-1} e^{-by_0(t-t_0)}[1 - e^{-b(t-t_0)}]^{y-y_0}.$$

Using this formula, expected value calculations will show that the covariance of y_i and y_j, $j \geqq i$, is given by

$$\sigma_{ij} = y_0 e^{b(t_j-t_0)}[e^{b(t_i-t_0)} - 1].$$

Under equal spacing as before, $t_0 = 0$ and $t_i = i\delta$; hence letting $z = e^{b\delta}$,

$$\sigma_{ij} = y_0 z^j(z^i - 1).$$

From this formula it follows that

$$\text{(13)} \qquad \sigma_{ij+m} = z^m \sigma_{ij} \quad \text{and} \quad \sigma_{jj} = \frac{z^j(z^j-1)}{z^i(z^i-1)}\sigma_{ii}$$

As a result, the covariance matrix S assumes the form

$$S = \begin{bmatrix} \sigma_{11} & z\sigma_{11} & \cdots & z^{n-1}\sigma_{11} \\ z\sigma_{11} & \sigma_{22} & \cdots & z^{n-2}\sigma_{22} \\ \vdots & \vdots & & \vdots \\ z^{n-1}\sigma_{11} & z^{n-2}\sigma_{22} & \cdots & \sigma_{nn} \end{bmatrix}$$

The second of formulas (13) enables this matrix to be expressed as the product of the following two matrices.

$$\begin{bmatrix} \dfrac{\sigma_{11}}{z-1} & 0 & \cdots & 0 \\ 0 & \dfrac{\sigma_{22}}{z^2-1} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \dfrac{\sigma_{nn}}{z^n-1} \end{bmatrix} \begin{bmatrix} z-1 & z^2-z & \cdots & z^n-z^{n-1} \\ z-1 & z^2-1 & \cdots & z^n-z^{n-2} \\ \vdots & \vdots & & \vdots \\ z-1 & z^2-1 & \cdots & z^n-1 \end{bmatrix}$$

Some rather lengthy calculations will show that the inverse matrix is given by

$$S^{-1} = \frac{1}{y_0(z-1)} \begin{bmatrix} z+1 & -z & \cdots & 0 & 0 \\ -1 & z+1 & \cdots & 0 & 0 \\ 0 & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & & & \\ 0 & 0 & \cdots & z+1 & -z \\ 0 & 0 & \cdots & -1 & 1 \end{bmatrix} \begin{bmatrix} \dfrac{1}{z} & 0 & \cdots & 0 \\ 0 & \dfrac{1}{z^2} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \dfrac{1}{z^n} \end{bmatrix}$$

Additional lengthy computations, similar to those employed in the preceding section, will show that (3) assumes the form

$$\frac{1}{\text{G.V.}(n,\delta)} = \frac{\delta^{k(k+1)} \mid C(n,z) \mid}{y_0^{k+1}(z-1)^k z^n}$$

where $C(n, z)$ is the matrix whose element in row $p + 1$ and column $q + 1$ is given by

$$c_{p+1\;q+1} = \frac{(2^p - 1^p)(2^q - 1^q)}{z} + \frac{(3^p - 2^p)(3^q - 2^q)}{z^2} + \cdots + \frac{(n^p - (n-1)^p)(n^q - (n-1)^q)}{z^{n-1}}.$$

For $p = 1$ or $q = 1$, $c_{p+1,q+1}$ is defined by $c_{11} = z^n / (z - 1) - 1$, and $c_{i1} = c_{1i} = n^{i+1} - 1$, $i > 1$.

When $n \to \infty$, the elements of this matrix exclusive of those in the first row and first column, converge to functions of z, for $z > 1$. Let

$$g_{pq}(z) = \lim_{n\to\infty} c_{p+1\;q+1}\,.$$

Since, for $z > 1$, c_{11} dominates c_{i1}, $i > 1$, the determinant $|c(n, z)|$ will possess the asymptotic value

$$\frac{z^n}{z-1}\begin{vmatrix} g_{11}(z) & \cdots & g_{1k}(z) \\ \vdots & & \vdots \\ g_{k1}(z) & \cdots & g_{kk}(z) \end{vmatrix}$$

For $k < 5$ it has been shown that the preceding determinant has the value

$$\frac{cz^{k(k-1)/2}}{(z-1)^{k^2}}$$

where c depends on k but not on z. Using these results the asymptotic value of the generalized variance is given by

$$\frac{1}{\text{G.V.}(n,\delta)} = \frac{\delta^{k(k+1)} c z^{k(k-1)/2}}{y_0^{k+1}(z-1)^{k^2+k+1}}. \tag{14}$$

From this result it is easily shown that

$$\lim_{n\to\infty} R_1 = \frac{(\sqrt{z}+1)^{(k^2+k+1)/(k+1)}}{2^k(\sqrt{z})^{(k_2-k)/(2k+2)}}.$$

Since (14) does not involve n, it follows that

$$\lim_{n\to\infty} R_2 = 1.$$

As a numerical illustration here, let $z = e^{b\delta} = 10/9$. This value implies that the correlation between y_1 and y_2 is approximately .7 and increases between

neighboring y values as one moves out on the axis. Calculations here yield the following limiting values for R_1.

k	1	2	3	4
$\lim_{n\to\infty} R_1$	1.47	1.32	1.25	1.20

These limiting values of R_1 show that some additional estimation information is gained by doubling the number of points in a fixed interval but that repeating the experiment yields considerably more information. The limiting value of R_2 would seem to indicate that no additional information is gained by extending the interval. This limiting result, however, is not realistic for small samples as will be seen in the next section.

9. Numerical results. Since the asymptotic measures of estimation efficiency obtained in the preceding sections may not be very realistic for small numbers of observations, some numerical computations were made with the assistance of high speed computing equipment. The values of $w = .64$ and $z = 10/9$ used previously were used in these computations. Values of $n = 5$ and $n = 10$ were chosen but only the results for $n = 10$ are given because some of the $n = 5$ values appeared questionable and because there were only moderate differences between the two sets of values. The limiting values of R_1 and R_2 are shown in parentheses adjacent to the computed values. In these computations, adjustments were made in the values of R_1 and R_2 to allow for the fact that doubling the number of points in an interval extends the total interval spanned by the points when the first point is located at $x = \delta$. These adjustments essentially kept the spanned interval unchanged. This was accomplished by replacing $\delta/2$ by $\delta(n - 1)/(2n - 1)$ in the denominator of R_1 and δ by $\delta(2n - 2)/(2n - 1)$ in the denominator of R_2.

k	Model (a)	Model (b)	Model (c)
1	1.90 (2)	1.03 (1.01)	1.43 (1.47)
2	1.81 (2)	1.02 (1.01)	1.25 (1.32)
3	1.72 (2)	1.02 (1.01)	1.19 (1.25)
4	1.64 (2)	1.03 (1.01)	1.18 (1.20)

k	Model (a)	Model (b)	Model (c)
1	3.80 (4)	2.91 (4)	1.83 (1)
2	7.24 (8)	5.01 (8)	1.85 (1)
3	13.76 (16)	8.94 (16)	3.21 (1)
4	26.24 (32)	16.41 (32)	6.15 (1)

It will be observed that the asymptotic values of R_2 are poor approximations for models (b) and (c). These results seem to indicate that in general one should always attempt to extend the range over which observations are to be taken as far as possible and the higher the degree of polynomial the greater is the advantage. They also seem to indicate that if the range can't be extended, it is considerably more efficient to replicate the experiment than double the number of observations, particularly if the variables are strongly correlated.

REFERENCES

[1] K. Smith, "On the standard deviations of adjusted and interpolated values of an observed polynomial function," *Biometrika*, Vol. 12 (1918), pp. 1–85.

[2] A. De La Garza, "Spacing of information in polynomial regression," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 123–30.

[3] G. Elfving, "Optimum allocation in linear regression theory," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 255–62.

[4] H. Chernoff, "Locally optimal designs for estimating parameters," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 586–602.

[5] H. E. Daniels. "A distribution free test for regression parameters," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 499–513.

[6] S. Ehrenfeld, "On the efficiency of experimental designs," *Ann. Math. Stat.*, Vol. 26 (1955), pp. 247–255.

[7] I. Schur, "Über die Verteilung der Wurzeln bei gewissen algebraischen Gleichungen mit ganzzahligen Koefficienten," *Math. Zeit.*, Vol. 1 (1918), p. 378.

[8] G. Prevost, *Tables et Formules de Fonctions Spheriques*, Gauthier-Villars, Paris, 1933.

[9] H. Cramér, *Mathematical Methods of Statistics*, Princeton University Press, 1946, p. 494.

[10] T. Muir, *Treatise on the Theory of Determinants*, Longmans, Green and Co. (1933), p. 440.

[11] P. G. Guest, "The spacing of observations in polynomial regression," *Ann. Math. Stat.*, Vol. 29 (1958), pp. 294–299.

ON THE GENERAL CANONICAL CORRELATION DISTRIBUTION

By A. G. Constantine and A. T. James

Division of Mathematical Statistics, C.S.I.R.O.[1]

1. Summary. The paper is divided into two parts:

A. An elementary derivation of Bartlett's results on the distribution of the canonical correlation coefficients using exterior differential forms. Briefly, our method consists of taking the original multivariate normal distribution, transforming to the canonical correlations and other variables, and then integrating out these extraneous variables.

B. A new method of calculating the conditional moments which appear in Bartlett's expansion of this distribution, based on the process of averaging over the orthogonal group. This method allows the calculation of moments of any order.

Part A

2. Introduction. Bartlett [1] obtained the general canonical correlation distribution as a multiple power series in the true canonical correlations ρ_i. In the case of more than one non-zero correlation ρ_i, the coefficients in this expansion depend on the conditional moments of the sample (ordinary) correlations s_i between the pairs of transformed variates representing the true canonical variates, when the sample canonical correlations r_i between the sample canonical variates are fixed.

Bartlett derived his results by a formal generalization of the argument used by Fisher [2] in calculating the distribution of the multiple correlation coefficient. We shall give a new proof of Bartlett's results in a concrete form more suitable for our purposes. Throughout this paper we shall use the concepts of exterior differential forms and alternating products of these forms. The definition and a discussion of these concepts may be found in James [6].

Consider a dependent vector variate with p components and an independent vector variate with $q \geqq p$ components. (Here the terms "dependent" and "independent" are to be understood in the regression sense.) If we take a sample with $n(\geqq p + q)$ degrees of freedom, we may represent it by the $p + q$ column vectors $\xi_1, \xi_2, \cdots, \xi_p$ and $\eta_1, \eta_2, \cdots, \eta_q$, each containing n components. The dependent vector is considered to be a normal variate, and we may distinguish two cases, according as the independent variate is assumed to be (a) a normal variate or (b) a set of fixed vectors in the sample space. In either case we may, without loss of generality, assume the ξ_i and η_j to be the canonical variates (see

Received September 23, 1957; revised March 1, 1958.

[1] Commonwealth Scientific and Industrial Research Organization, University of Adelaide, South Australia.

Hotelling [4]). This means that in case (a) the n components of each vector are standard normal variates with the joint distribution

$$\prod_{i=1}^{p}\left\{(2\pi)^{-n}(1-\rho_i^2)^{-n/2}\exp\left[\frac{-(\xi_i'\xi_i-2\rho_i\xi_i'\eta_i+\eta_i'\eta_i)}{2(1-\rho_i^2)}\right]\prod_{v=1}^{n}d\xi_{vi}\,d\eta_{vi}\right\} \cdot \prod_{j=p+1}^{q}\left\{(2\pi)^{-n/2}\exp\left[-\eta_j'\eta_j/2\right]\prod_{v=1}^{n}d\eta_{vj}\right\}. \tag{2.1}$$

In case (b), the non-central means case, we may assume the components of the ξ_i to be independently distributed with unit variance, and the η_j to be vectors lying along the first q co-ordinate axes of the sample space. $\eta_1, \cdots, \eta_p$ may also be identified with the mean vectors of $\xi_1, \cdots, \xi_p$. The joint distribution of the ξ_{vi} is therefore

$$\prod_{i=1}^{p}\left\{(2\pi)^{-n/2}\exp\left[-(\xi_i'\xi_i-2\xi_i'\eta_i+\eta_i'\eta_i)/2\right]\prod_{v=1}^{n}d\xi_{vi}\right\}. \tag{2.2}$$

We denote sample correlations between ξ_i and η_i by s_i and the sample canonical correlations between the sample canonical variates by r_i. The r_i may also be interpreted as the cosines of the critical angles between the two planes spanned by $x_1, \cdots, x_p$ and $y_1, \cdots, y_q$ respectively, where the x_i and y_j are the sample canonical variates. The distribution of the r_i for each of the two cases mentioned above will be derived in sections 3 and 4 respectively.

3. Distribution of the canonical correlation coefficients. Our starting point is the distribution (2.1). The distribution of the canonical correlations r_i will be obtained by expressing this distribution in terms of the r_i and other variables and integrating over the ranges of the latter. First of all, let us dispose of the lengths of the vectors ξ_i and η_j.

Put $\xi_i = \tau_i w_i$ and $\eta_j = \sigma_j z_j$ where τ_i and σ_j are the unit vectors along ξ_i and η_j respectively, and w_i and z_j are their lengths. Then

$$\prod_{v=1}^{n} d\xi_{vi} = w_i^{n-1}\,dw_i\,dS(\tau_i) \tag{3.1}$$

where $dS(\tau_i)$ is the element of area on the unit sphere in n-space. With an analogous expression for $\prod d\eta_{vj}$ the distribution (2.1) becomes

$$\begin{aligned}\prod_{i=1}^{p}&\left\{\frac{1}{2^{n-2}(1-\rho_i^2)^{n/2}[\Gamma(n/2)]^2}\right.\\ &\left.\cdot\exp\left[-\frac{1}{2(1-\rho_i^2)}(w_i^2+z_i^2-2\rho_i s_i w_i z_i)\right](w_iz_i)^{n-1}\,dw_i\,dz_i\right\}\\ &\times\prod_{j=p+1}^{q}\frac{1}{2^{(n-2)/2}\Gamma(n/2)}\exp\left[-\tfrac{1}{2}z_j^2\right]z_j^{n-1}\,dz_j\prod_{i=1}^{p}\frac{\Gamma(n/2)}{2\pi^{n/2}}\,dS(\tau_i)\\ &\times\prod_{j=1}^{q}\frac{\Gamma(n/2)}{2\pi^{n/2}}\,dS(\sigma_j),\end{aligned} \tag{3.2}$$

where $s_i = \tau_i'\sigma_i$ (see section 2). The constants have been split up to make the latter factors probability distributions.

The integrals of the factors containing z_j for $j = p + 1, \cdots, q$ are obviously unity. Furthermore, by expanding the factor $\exp[(1 - \rho_i^2)^{-1}\rho_i s_i w_i z_i]$ in a power series and integrating term-by-term with respect to w_i and z_i $(i = 1, \cdots, p)$ we obtain

$$\begin{aligned}(3.3)\qquad &\int_0^\infty\int_0^\infty \frac{1}{2^{n-2}(1-\rho_i^2)^{n/2}[\Gamma(n/2)]^2} \\ &\quad\cdot \exp[-(w_i^2 + z_i^2 - 2\rho_i s_i w_i z_i)/2(1-\rho_i^2)](w_i z_i)^{n-1}\,dw_i\,dz_i \\ &\qquad = (1-\rho_i^2)^{n/2}{}_2F_1(n/2, n/2; 1/2; \rho_i^2 s_i^2) + \text{an odd function of } s_i,\end{aligned}$$

where ${}_2F_1$ is the Gaussian hypergeometric function. Later on, we shall see that the odd function of s_i vanishes in the subsequent integrations.

The next step is to express the unit column vectors τ_i and σ_j in terms of the canonical correlations r_i and the vectors x_i and y_j which determine these correlations. Let $\mathfrak{p}$ and $\mathfrak{q}$ be the p-plane and the q-plane spanned by the vectors τ_i and σ_j in n-space. Then $\mathfrak{p}$ and $\mathfrak{q}$ determine almost certainly (i.e. with probability 1) the orthonormal vectors x_i and y_i $(i = 1, \cdots, p)$ which make the critical angles between the planes, i.e. such that $x_i'y_i = r_i$, $x_i'y_j = 0$ $(i \neq j)$. Let further vectors $y_{p+1}, \cdots, y_q$ be defined as functions of $\mathfrak{p}$ and $\mathfrak{q}$ to complete an orthonormal set spanning $\mathfrak{q}$. T, Σ, X, Y will denote the matrices composed of the column vectors τ_i, σ_j, x_i, y_j, respectively. It follows that $X'X = I_p$, $Y'Y = I_q$ and $X'Y = [R \,\vdots\, 0]$ where R is the diagonal matrix with the r_i down the main diagonal. Furthermore we may write

$$(3.4)\qquad T = XA, \qquad \Sigma = YB$$

where A is a $p \times p$ and B is a $q \times q$ matrix. Then $T'T = A'A$ and $\Sigma'\Sigma = B'B$. The matrices A and B are subject only to the restriction that all their columns α_i and β_j are of unit length.

We now substitute for $\prod dS(\tau_i)$ and $\prod dS(\sigma_j)$ in (3.2), using the transformations (3.4). To avoid interrupting the continuity of the argument we shall, for the moment, only give the results of the substitution, and defer the proof until section 5. We have then from (5.4)

$$(3.5)\qquad \prod_{i=1}^p dS(\tau_i) = |A'A|^{(n-p)/2}\prod_{i=1}^p dS(\alpha_i)\,d\mathfrak{p} + {}^*(dX)\,d\mathfrak{p}$$

where $dS(\alpha_i)$ is the element of area on the unit sphere in p-space and $d\mathfrak{p}$ is the differential form representing the invariant measure on the Grassmann manifold of p-planes in n-space. The symbol ${}^*(dX)$ stands for certain differentials involving the elements of X, which, when subsequently multiplied by other differentials, will vanish. Similarly

$$(3.6)\qquad \prod_{j=1}^q dS(\sigma_j) = |B'B|^{(n-q)/2}\prod_{j=1}^q dS(\beta_j)\,d\mathfrak{q} + {}^*(dY)\,d\mathfrak{q}$$

where $dS(\beta_j)$ is the element of area on the unit sphere in q-space. Multiplying (3.5) and (3.6) we obtain

$$\prod_{i=1}^{p} dS(\tau_i) \prod_{j=1}^{q} dS(\sigma_j) = | A'A |^{(n-p)/2} \prod_{i=1}^{p} dS(\alpha_i) \, | B'B |^{(n-q)/2} \prod_{j=1}^{q} dS(\beta_j) \, d\mathfrak{p} \, d\mathfrak{q}. \tag{3.7}$$

The terms containing $^*(dX)$ and $^*(dY)$ vanish when multiplied by $d\mathfrak{p}\, d\mathfrak{q}$, since $d\mathfrak{p}\, d\mathfrak{q}$ is of maximum degree in $\mathfrak{p}$ and $\mathfrak{q}$ and X and Y are functions of $\mathfrak{p}$ and $\mathfrak{q}$.

The differential form $d\mathfrak{p}\, d\mathfrak{q}$ may now be expressed in terms of the r_i and other variables. Integration with respect to these latter variables yields

$$K_p K_q \phi(r_i \mid \rho_i = 0)$$

where K_p and K_q are the normalising constants of the differential forms $d\mathfrak{p}$ and $d\mathfrak{q}$ respectively, and $\phi(r_i \mid \rho_i = 0)$ is the null distribution of the r_i (see James [6]):

$$\phi(r_i \mid \rho_i = 0) = C \prod_{i=1}^{p} \{(r_i^2)^{(q-p-1)/2}(1 - r_i^2)^{(n-q-p-1)/2}\} \prod_{i<j} (r_i^2 - r_j^2) \prod_{i=1}^{p} dr_i^2$$

and

$$C = \pi^{p/2} \prod_{i=0}^{p-1} \left\{ \Gamma\left(\frac{n-i}{2}\right) \Big/ \Gamma\left(\frac{p-i}{2}\right) \Gamma\left(\frac{q-1}{2}\right) \Gamma\left(\frac{n-q-i}{2}\right) \right\}.$$

This distribution was first derived by Fisher [3], Hsu [5] and Roy [8]. The values of K_p and K_q are given by

$$K_v = \prod_{i=1}^{v} \frac{G(n-i+1)}{G(i)}, \qquad G(i) = \frac{2\pi^{i/2}}{\Gamma(i/2)}, \qquad v = p, q.$$

After this integration, the right hand side of (3.7) becomes

$$K_p \mid A'A \mid^{(n-p)/2} \prod dS(\alpha_i) K_q \mid B'B \mid^{(n-q)/2} \prod dS(\beta_j) \cdot \phi(r_i \mid \rho_i = 0), \tag{3.8}$$

showing that A, B and the r_i are independently distributed.

Substituting (3.3) and (3.8) in (3.2), we may write the distribution of the r_i as

$$\begin{aligned} \int_A \int_B \prod_{i=1}^{p} \{(1 - \rho_i^2)^{n/2} {}_2F_1(n/2, n/2; 1/2; \rho_i^2 s_i^2)\} k_p \mid A'A \mid^{(n-p)/2} \\ \cdot \prod_{i=1}^{p} dS(\alpha_i) k_q \mid B'B \mid^{(n-q)/2} \prod_{j=1}^{q} dS(\beta_j) \phi(r_i \mid \rho_i = 0), \end{aligned} \tag{3.9}$$

together with the relation

$$s_i = \tau_i' \sigma_i = \alpha_{1i}\beta_{1i}r_1 + \alpha_{2i}\beta_{2i}r_{2i} + \cdots + \alpha_{pi}\beta_{pi}r_p . \tag{3.10}$$

The normalising constants k_p and k_q for the distribution of A and B are given by

$$k_v = \prod_{i=1}^{v} \frac{G(n-i+1)}{G(n)G(i)}, \qquad v = p, q. \tag{3.11}$$

In view of equation (3.10) we may now identify our distribution (3.9) with Bartlett's distribution, [1], equations (8) and (10).

If the hypergeometric functions are expanded as power series and multiplied together, the function multiplying $\phi(r_i \mid \rho_i = 0)$ is seen to be a multiple power series in the ρ_i whose coefficients depend on the expectations of monomials in the s_i with respect to the distribution

$$(3.12) \qquad k_p \mid A'A \mid^{(n-p)/2} dS(\alpha_1) \cdots dS(\alpha_p)$$

of A and a similar distribution of B.

So far we have ignored the odd function of s_i appearing in the integral (3.3). However, any odd function $f(s_i)$ of s_i will have zero expectation. In fact, putting $-\alpha_i$ instead of α_i does not alter the distribution (3.12) of A, but changes s_i to $-s_i$ in view of (3.10). Therefore,

$$E[f(s_i)] = E[f(-s_i)] = E[-f(s_i)] = -E[f(s_i)]$$

and so $E[f(s_i)] = 0$. It is sufficient, therefore, to compute only moments of the form $\mu(t_1, t_2, \cdots, t_p) = E\{(s_1^2)^{t_1}(s_2^2)^{t_2} \cdots (s_p^2)^{t_p}\}$ where the expectations are taken with respect to the distributions of A and B and the r_i are held fixed. Furthermore, if we substitute in (3.9) for s_i using (3.10), the calculations are reduced to finding the moments of the α_{ij} and β_{ij}, two independent sets of variates.

Theoretically these moments could be found directly from the distributions of A and B. However, as Bartlett pointed out, this method is too difficult algebraically to be of much use, except in the case of only one non-zero ρ_i. Bartlett indicated a method whereby moments of the form $\mu(t_1, t_2)$ could be calculated, and also calculated $\mu(1, 1, 1)$ by employing various relations connecting the α-moments (see section 10). Again, both of these methods led to awkward algebra and had to be abandoned for moments of higher order, though Bartlett was able to compute $\mu(1, 1)$ $\mu(2, 1)$, $\mu(2, 2)$ and $\mu(3, 1)$. In part B of this paper we shall present a method enabling moments of any order to be computed, and shall complete the tabulation of moments up to the fourth order with $\mu(2, 1, 1)$ and $\mu(1, 1, 1, 1)$.

4. The non-central means case. Let $\mathfrak{p}$ be the random plane spanned by the vectors $\xi_1, \cdots, \xi_p$ and $\mathfrak{q}$ the fixed plane spanned by $\eta_1, \eta_2, \cdots, \eta_q$. As we saw in section 2, we may assume that the $\xi_1, \cdots, \xi_p$ are independently distributed and their components ξ_{vi} have the distribution

$$(4.1) \qquad \prod_{i=1}^{p} (2\pi)^{-n/2} \exp\left[-(\xi_i' \xi_i - 2\xi_i' \eta_i + \eta_i' \eta_i)/2\right] \prod_{v=1}^{n} d\xi_{vi}.$$

Furthermore, the η_j $(j = 1, \cdots, q)$ may be taken as vectors lying along the first q co-ordinate axes in the sample space and thus having only one non-zero component each, say μ_j in the jth position.

Putting $\xi_i = \tau_i w_i$ as before, (4.1) becomes

$$\prod_{i=1}^{p} \frac{1}{2^{(n-2)/2}\Gamma(n/2)} \exp\left[-(w_i^2 - 2\mu_i s_i w_i + \mu_i^2)/2\right] w_i^{n-1}\, dw_i \times \prod_{i=1}^{p} \frac{\Gamma(n/2)}{2\pi^{n/2}}\, dS(\tau_i), \tag{4.2}$$

where $s_i = \tau_{ii}$. The integral with respect to w_i of the ith factor in the first product of (4.2) is ${}_1F_1(n/2; 1/2; \mu_i^2 s_i^2/2)e^{-\mu_i^2/2}$ + an odd function of s_i. This odd function will again vanish in subsequent integrations and may be ignored from now on.

Let X be the $n \times p$ matrix whose columns are the orthonormal vectors $x_1, x_2, \cdots, x_p$ spanning $\mathfrak{p}$ and which make the critical angles with $\mathfrak{q}$. The τ_i may be expressed as linear combinations of the x_i by putting

$$T = XA. \tag{4.3}$$

Since $X'X = I_p$ we have $T'T = A'A$. From section 5, (5.4), it follows that

$$\prod_{i=1}^{p} dS(\tau_i) = |A'A|^{(n-p)/2} \prod_{i=1}^{p} dS(\alpha_i)\, d\mathfrak{p}, \tag{4.4}$$

the differential form $^*(dX)\, d\mathfrak{p}$ vanishing since X and $\mathfrak{p}$ are functions of each other.

To express $\mathfrak{p}$ in terms of the r_i, we partition X as follows:

$$X = \begin{bmatrix} Y \\ \cdots \\ Z \end{bmatrix} \tag{4.5}$$

where Y is a $q \times p$ matrix and Z is an $(n - q) \times p$ matrix. The vector $\begin{bmatrix} y_i \\ \cdots \\ 0 \end{bmatrix}$ in $\mathfrak{q}$ makes the ith critical angle with x_i in $\mathfrak{p}$. Let β_i and $\gamma_i (i = 1, \cdots, p)$ be the unit vectors along y_i and z_i, then according to [6], equation (7.10),

$$y_i = \beta_i r_i, \qquad z_i = \gamma_i \sqrt{1 - r_i^2} \tag{4.6}$$

and

$$d\mathfrak{p} = K_p \frac{1}{\prod_{i=1}^{p} G(q - i + 1)} \cdot dV(\beta) \frac{1}{\prod_{i=1}^{p} G(n - q - i + 1)} dV(\gamma)\phi(r_i \mid \rho_i = 0) \tag{4.7}$$

where K_p and $G(i)$ are defined in section 3, and $dV(\beta)$ and $dV(\gamma)$ are the invariant measures on the Stiefel manifolds of p-frames $(\beta_1, \cdots, \beta_p)$ in q-space and p-frames $(\gamma_1, \cdots, \gamma_p)$ in $(n - q)$-space. The constant has been split up to nor-

malise these invariant measures. If we choose $q - p$ orthonormal vectors $\beta_{p+1}, \cdots, \beta_q$ orthogonal to $\beta_1, \cdots, \beta_p$ we may express $dV(\beta)$ as

$$dV(\beta) = \prod_{i<j}^{p} \beta_j' \, d\beta_i \prod_{j=p+1}^{q} \prod_{i=1}^{p} \beta_j' \, d\beta_i . \tag{4.8}$$

Also,

$$s_i = \tau_{ii} = \sum_{j=1}^{p} x_{ij}\alpha_{ji} = \sum_{j=1}^{p} \beta_{ij} r_j \alpha_{ji}$$

If we please, we may replace β_{ij} by β_{ji} since they have the same distribution. Integrating (4.7) with respect to γ, substituting in (4.4) and then in (4.2), we obtain the distribution of the r_i as

$$\begin{aligned} &\int_A \int_B \prod_{i=1}^{p} {}_1F_1(n/2;\, 1/2;\, \tfrac{1}{2}\mu_i^2 s_i^2) e^{-\mu_i^2/2} k_p \,|\, A'A \,|^{(n-p)/2} \prod_{i=1}^{p} dS(\alpha_i) \\ &\qquad \cdot \frac{1}{\prod_{i=1}^{p} G(q - i + 1)} \beta_j' \, d\beta_i \prod_{j=p+1}^{q} \prod_{i=1}^{p} \beta_j' \, d\beta_i \phi(r_i \,|\, \rho_i = 0(\end{aligned} \tag{4.9}$$

where

$$s_i = \alpha_{1i}\beta_{1i} r_1 + \cdots + \alpha_{pi}\beta_{pi} r_p . \tag{4.10}$$

We notice that the distribution of A is identical with its distribution in the previous case, but now the distribution of B is the invariant distribution on a Stiefel manifold and is independent of n. However, A and B are still independent.

5. Distribution of the co-ordinates of random vectors in a random plane. In relation to the rest of the paper, the purpose of this section is to derive equation (3.5) and a result at the end of section 7. However, the results will be more interesting and intelligible if discussed in terms of probabilities.

$\tau_1, \cdots, \tau_p$ are invariantly distributed unit vectors in n-space, which we write as the columns of an $n \times p$ matrix T. $\mathfrak{p}$ is the plane spanned by the τ_i. We define in $\mathfrak{p}$ a reference set of orthonormal vectors, which we write as the columns of an $n \times p$ matrix X. Thus X is a function of $\mathfrak{p}$ and

$$X'X = I_p . \tag{5.1}$$

Let the column α_i of the $p \times p$ matrix A be the co-ordinates of τ_i relative to the reference set X:

$$T = XA. \tag{5.2}$$

We shall show that $\mathfrak{p}$ is invariantly distributed and that A is independently distributed with density proportional to

$$|\, A'A \,|^{(n-p)/2} \prod_{i=1}^{p} dS(\alpha_i). \tag{5.3}$$

These results are implicit in Bartlett [1]. They follow from the lemma which we shall now state and prove. For the application in section 3 we shall have to generalise the situation slightly to include the case when the reference set X is not necessarily a function of $\mathfrak{p}$ alone.

LEMMA. *If T is an $n \times p$ matrix whose columns τ_i are unit vectors, and X and A are $n \times p$ and $p \times p$ matrices satisfying* (5.1) *and* (5.2), *then*

$$\prod_{i=1}^{p} dS(\tau_i) = |A'A|^{(n-p)/2} \prod_{i=1}^{p} dS(\alpha_i)\, d\mathfrak{p} + {}^*(dX)\, d\mathfrak{p} \tag{5.4}$$

where $^*(dX)$ is a differential form in X and A, every term of which is of at least first degree in dX. If X is a function of $\mathfrak{p}$ alone, then $^*(dX)\, d\mathfrak{p} = 0$.

PROOF. Selecting a single column from the matrix equation (5.2) we have

$$\tau_i = X\alpha_i . \tag{5.5}$$

Differentiating:

$$d\tau_i = dX\alpha_i + X\, d\alpha_i . \tag{5.6}$$

As the differential form for $dS(\alpha_i)$ will be required, we introduce $p - 1$ orthonormal column vectors in p-space orthogonal to α_i. Let C_i be the $p \times p - 1$ matrix with them as columns. Then $dS(\alpha_i)$ is the alternating product of the elements in the vector $C_i'\, d\alpha_i$.

The differential form for $dS(\tau_i)$ requires $n - 1$ orthonormal vectors orthogonal to τ_i. The columns of the matrix XC_i provide $p - 1$ of them, since $C_i'X'\tau_i = C_i'X'X\alpha_i = C_i'\alpha_i = 0$. Choose the remaining $n - p$ orthonormal vectors orthogonal to the plane $\mathfrak{p}$ and let them be columns of an $n \times (n - p)$ matrix B, which is to be a function merely of p.

Premultiply (5.6) by the transpose of the partitioned matrix $[XC_i \,\vdots\, B]$:

$$\begin{bmatrix} C_i' X\, d\tau_i \\ B'\, d\tau_i \end{bmatrix} = \begin{bmatrix} C_i' X'\, dX\alpha_i + C_i'\, d\alpha_i \\ B'\, dX\alpha_i \end{bmatrix}. \tag{5.7}$$

Then, the alternating product of the differentials of the vector on the left is $dS(\tau_i)$ and hence the product of all of these for $i = 1, \cdots, p$ is the density on the left-hand side of (5.4).

The alternating product of all the differentials in the right-hand side of (5.7) for $i = 1, \cdots, p$ will give the density in the new co-ordinates. Let us deal with the vector differentials $B'\, dX\alpha_i$ first. These p vector differentials, corresponding to $i = 1, \cdots, p$, comprise the columns of the matrix $B'\, dXA$, of whose elements we therefore want the alternating product. The alternating product of the elements of a *row* of this matrix is $|A|$ times the product of the row of the elements of $B'\, dX$. There being $n - p$ rows in $B'\, dXA$, the alternating product of all its elements is then $|A|^{n-p} \prod_j \prod_j b_j'\, dx_i$. The differential form

$$\prod_j \prod_i b_j'\, dx_i$$

is the invariant measure, $d\mathfrak{p}$, on the Grassmann manifold, i.e. the uniform distribution of a p-plane in n-space (see [6]).

As the differential forms represent probability densities and must therefore be positive, we replace $| A |$ by its modulus $| A'A |^{1/2}$.

The product of the elements of the vector $C_i' \, d\alpha_i$ is $dS(\alpha_i)$. All the products involving an element of $C_i'X' \, dX\alpha_i$ we lump together in the symbol $* \, dX$. Collecting all factors we obtain (5.4). Q.E.D.

We conclude with a result that we shall need in section 7. From (5.1) and (5.2) we have $T'T = A'A$. Hence, if A has the distribution (5.3) then the moments of $A'A$ are the same as the moments of $T'T$ where T has the distribution $\prod dS(\tau_i)$.

PART B

6. Introduction. In this part of the paper we shall be concerned with the problem of calculating the conditional moments

$$\mu(t_1, t_2, \cdots, t_p) = E[(s_1^2)^{t_1}(s_2^2)^{t_2} \cdots (s_p^2)^{t_p}]$$

required for the expansion of the distribution of the canonical correlations r_i.

Recalling the results of sections 3 and 4, we saw that the expectations of monomials in the s_i^2 could be replaced by the expectations of monomials $m(A, B)$ in $\alpha_{ij}\beta_{ij}$ in view of the relation

$$s_i = \alpha_{1i}\beta_{1i}r_1 + \cdots + \alpha_{pi}\beta_{pi}r_p . \tag{6.1}$$

Furthermore, since $A = (\alpha_{ij})$ is distributed independently of $B = (\beta_{ij})$,

$$E[m(A, B)] = E[m(A)] \, E[m(B)]$$

where $m(A)$ and $m(B)$ are monomials in the elements of A and B respectively. Considering case (a) for the moment, we saw that the distributions of A and B were

$$k_p \, | A'A |^{(n-p)/2} \, dS(\alpha_1) \cdots dS(\alpha_p), \tag{6.2}$$

and

$$k_q \, | B'B |^{(n-q)/2} \, dS(\beta_1) \cdots dS(\beta_q)$$

respectively. Consequently, $E[m(B)]$ may be obtained from $E[m(A)]$ by simply replacing p with q.

In case (b), though the distribution of A is still given by (6.2), the distribution of B is given by (4.8), the invariant distribution on the Stiefel manifold of p-frames in q-space. We notice, however, that if we let $n \to \infty$ in case (a), then the set of random vectors $(\beta_1, \cdots, \beta_p)$ becomes a rigid p-frame, and this, of course, is exactly the situation in case (b). Hence the β-moments may be obtained from those in case (a) by letting $n \to \infty$. To summarise, then, it is sufficient to compute only the moments of the distribution (6.2).

To compute these moments by direct integration is obviously going to lead

to involved algebra. However, by first averaging the monomials $m(A)$ over the orthogonal group we can considerably simplify the problem. Before proceeding further we shall briefly discuss this important process.

7. Average over the orthogonal group. The process $\mathfrak{M}$ of averaging over a group is a linear process whereby a function, defined on a space on which a group of transformations acts, is changed into a function invariant under the group. In particular, we consider the group $\mathfrak{H}$ of all orthogonal matrices H, and a matrix $A = (\alpha_{ij})$ which is transformed by the elements of $\mathfrak{H}$:

$$A \rightarrow HA \tag{7.1}$$

If $f(A)$ is a function of the elements of A, then

$$\mathfrak{M}f(A) = \int_{\mathfrak{H}} f(H^{-1}A)\, dV(H)$$

is a function invariant under the transformations (7.1). $V(H)$ is the invariant measure on the orthogonal group, normalised so that $V(\mathfrak{H}) = 1$. $\mathfrak{M}f$ is called the *average or mean value* of the function over the group. (This definition of "mean value" should not be confused with the usual statistical definition.) Since $\mathfrak{M}f$ is invariant under the orthogonal group, it must be expressible as a function of the basic invariants $\alpha_i'\alpha_j$ (see Weyl [9], pp. 52–6).

We wish to calculate the expectations of monomials $m(A)$ in the elements of A. Since the distribution (6.2) is invariant under the transformations (7.1), $E[m(A)] = E[m(H^{-1}A)]$, and hence it follows that

$$E[m(A)] = \int E[m(A)]\, dV(H) = \int E[m(H^{-1}A)]\, dV(H)$$

$$= E \int m(H^{-1}A)\, dV(H) = E[\mathfrak{M}m(A)].$$

In section 8 we shall show how to calculate $\mathfrak{M}m(A)$.

However, assuming for the moment that this has been done, we see that the problem has been reduced to the evaluation of the expectations of certain invariant functions $\phi(A'A)$, say. At this point it should be noted that the problem of the β-moments in case (b) has been completely solved. For, if we let $n \rightarrow \infty$, then $B'B = I$ with probability 1, and hence $E[m(B)] = \phi(I)$. $E[m(B)]$ can be then evaluated by the method given in James [7], pp. 374–5. However, since we require the β-moments for case (a), we may as well compute those for case (b) by letting $n \rightarrow \infty$ in the former moments, as indicated in section 6.

For the α-moments (and the β-moments for case (a)), we still have to evaluate the expectations of the invariant functions. In section 5 we have shown that the $\alpha_i'\alpha_j$ have the same distribution as quantities $\tau_i'\tau_j$ where $\tau_1, \cdots, \tau_p$ are independently uniformly distributed unit vectors in n-space. Finally, then, there remains the calculation of the moments of the $\tau_i'\tau_j$. This will be accomplished in section 9.

8. Calculation of $\mathfrak{M}m(A)$. In section 7 it was shown that

$$E[m(A)] = E[\mathfrak{M}m(A)] = E[\phi(A'A)].$$

In this section we shall show how to evaluate $\mathfrak{M}m(A)$.

Let

$$m(A) = \alpha_{i_1 j_1}^{k_1} \alpha_{i_2 j_2}^{k_2} \cdots \tag{8.1}$$

denote a monomial in the α_{ij}. Then if C is an arbitrary $p \times p$ matrix, the expansion of the function exp (tr $C'A$) contains every monomial (8.1) multiplied by the same monomial $m(C)$ in the corresponding elements of C, and divided by $k_1 ! \, k_2 ! \cdots$. James [7] has shown that $\mathfrak{M}$ exp (tr $C'A$) can be expanded as a multiple power series in the elementary symmetric functions $z_1, z_2, \cdots, z_p$ of the latent roots of $C'CA'A$. Thus, if $\lambda_1, \cdots, \lambda_p$ are the latent roots of $C'CA'A$, then

$$z_1 = \sum \lambda_i = \text{tr } C'CA'A,$$

$$z_2 = \sum_{i<j} \lambda_i \lambda_j = \text{sum of principal 2nd order minors of } C'CA'A, \text{ etc.},$$

and

$$\begin{aligned} \mathfrak{M} \exp (\text{tr } C'A) = 1 &+ \frac{z_1}{2p} + \frac{z_1^2}{8p(p+2)} + \frac{z_2}{2p(p+2)(p-1)} \\ &+ \frac{z_1^3}{8 \cdot 3! p(p+2)(p+4)} + \frac{z_1 z_2}{4p(p+2)(p+4)(p-1)} \\ &+ \frac{z_3}{p(p+2)(p+4)(p-1)(p-2)} + \frac{z_1^4}{2^4 4! p(p+2)(p+4)(p+6)} \\ &+ \frac{z_1^2 z_2}{16p(p+2)(p+4)(p+6)(p-1)} \\ &+ \frac{z_2^2}{8p(p+2)(p+4)(p+6)(p-1)(p+1)} \\ &+ \frac{(p+2) z_1 z_3}{2p(p+2)(p+4)(p+6)(p-1)(p+1)(p-2)} \\ &+ \frac{(5p+6) z_4}{2p(p+2)(p+4)(p+6)(p-1)(p+1)(p-2)(p-3)} + \cdots . \end{aligned} \tag{8.2}$$

Hence, $\mathfrak{M}m(A)$ can be found by equating the coefficients of $m(C)$ on both sides of (8.2).

If we write $A'A$ in the form

$$\begin{bmatrix} 1 & \alpha_1'\alpha_2 & \alpha_1'\alpha_3 & \cdots & \alpha_1'\alpha_p \\ \alpha_1'\alpha_2 & 1 & \alpha_2'\alpha_3 & \cdots & \cdot \\ \alpha_1'\alpha_3 & \alpha_2'\alpha_3 & 1 & & \cdot \\ \vdots & \vdots & \vdots & & \vdots \\ \alpha_1'\alpha_p & \alpha_2'\alpha_p & & & 1 \end{bmatrix} \tag{8.3}$$

we see that $\mathfrak{M}m(A)$ will be a linear combination of monomials in the invariants $\alpha_i'\alpha_j$. The expansion (8.2) is sufficient to compute all conditional moments up to order 4. If higher moments are required, further terms can be added to (8.2) by the use of recurrence relations derived from the differential equations given in James [7].

9. Calculation of the moments of the invariants. We are given that $\tau_1, \tau_2, \cdots, \tau_p$ are independently uniformly distributed column vectors in n-space, and we require the expectations of monomials in $\tau_i'\tau_j$. If a monomial in $\tau_i'\tau_j$ were expanded as a sum of monomials in the τ_{ij}, the expectations of each of these could be calculated and summed. However, the expansions would become very complicated. They can be avoided by the following method, which is an extension of an idea due to Bartlett [1] p. 13.

Let $e_1, e_2, \cdots, e_p$ be the unit vectors along the first p coordinate axes. Then the joint distribution of $\tau_1, \cdots, \tau_p$ is the same as that of $A_1e_1, A_2e_2, \cdots, A_pe_p$, where the A_i are random orthogonal matrices independently and invariantly distributed (see James [6]). Furthermore, the invariant functions will not be altered if they are calculated from the vectors $e_1, A_1'A_2e_2, \cdots, A_1'A_pe_p$. These vectors have the same distribution as $e_1, A_2e_2, \cdots, A_pe_p$ since $A_1'A_2, \cdots, A_1'A_p$ are still independently invariantly distributed. Again, if $A_2 = (a_{ij})$, say, the invariant functions will not be altered if we replace the vectors by

$$e_1, B_2'A_2e_2, \cdots, B_2'A_pe_p$$

where

$$B_2 = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & a_{22}/b_{22} & * & \cdots & \\ 0 & a_{32}/b_{22} & * & \cdots & \\ \cdot & \cdot & & \cdots & \\ \cdot & \cdot & & \cdots & \\ \cdot & \cdot & & \cdots & \\ 0 & a_{n2}/b_{22} & & \cdots & \end{bmatrix},$$

$b_{22}^2 = 1 - a_{12}^2 = a_{22}^2 + \cdots + a_{n2}^2$, and the remaining elements are chosen so that B_2 is orthogonal. Clearly,

$$B_2'A_2e_2 = \begin{bmatrix} a_{12} \\ b_{22} \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

Since the matrices $B_2'A_3, \cdots, B_2'A_p$ are still independently invariantly distributed we may replace the vectors by

$$e_1, B_2'A_2e_2, A_3e_3, \cdots, A_pe_p.$$

Proceeding in this way we see that we obtain the same values for the expectations of the invariants if we replace $\tau_1, \tau_2, \cdots, \tau_p$ by

$$\text{(9.1)} \qquad \begin{bmatrix} 1 \\ 0 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{bmatrix}, \quad \begin{bmatrix} a_{12} \\ b_{22} \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{bmatrix}, \quad \begin{bmatrix} a_{13} \\ a_{23} \\ b_{33} \\ 0 \\ \cdot \\ \cdot \\ 0 \end{bmatrix}, \cdots.$$

(To avoid introducing further notation, we have denoted the third column of A_3 by the elements $a_{13}, a_{23}, \cdots, a_{n3}$, those of the fourth column of A_4 by $a_{14}, a_{24}, \cdots, a_{n4}$ etc. Then $b_{33}^2 = 1 - a_{13}^2 - a_{23}^2$, $b_{44}^2 = 1 - a_{14}^2 - a_{24}^2 - a_{34}^2$, etc.)

EXAMPLE 1. As an example let us evaluate

$$E[(\alpha_1'\alpha_2)(\alpha_2'\alpha_3)(\alpha_3'\alpha_4)(\alpha_1'\alpha_4)] = E[(\tau_1'\tau_2)(\tau_2'\tau_3)(\tau_3'\tau_4)(\tau_1'\tau_4)].$$

Substituting from (9.1), this expectation is equal to

$$\text{(9.2)} \qquad E[a_{12}(a_{12}a_{13} + b_{22}a_{23})(a_{13}a_{14} + a_{23}a_{24} + b_{33}a_{34})a_{14}].$$

Now any monomial in the a_{ij}, b_{ii} containing an odd power has zero expectation since the distribution is unaltered if we replace a_{ij} by $-a_{ij}$ or b_{ii} by $-b_{ii}$. Hence, (9.2) reduces to $E(a_{12}^2a_{13}^2a_{14}^2)$. a_2, a_3 and a_4 are independently uniformly distributed unit vectors, and hence $E(a_{12}^2) = E(a_{13}^2) = E(a_{14}^2) = 1/n$. Therefore,

$$E[(\alpha_1'\alpha_2)(\alpha_2'\alpha_3)(\alpha_3'\alpha_4)(\alpha_1'\alpha_4)] = 1/n^3.$$

EXAMPLE 2. $E(\Delta)$ where $\Delta = |A'A|$.
Put

$$C = \begin{bmatrix} 1 & a_{12} & a_{13} & \cdots \\ 0 & b_{22} & a_{23} & \cdots \\ 0 & 0 & b_{33} & \cdots \\ \vdots & \vdots & \vdots & \end{bmatrix}.$$

Then $\Delta = |C'C| = |C|^2$, and

$$E(\Delta) = E(1 \cdot b_{22}^2 \cdot b_{33}^2 \cdots b_{pp}^2)$$

$$= 1 \cdot \frac{n-1}{n} \cdot \frac{n-2}{n} \cdots \frac{n-p+1}{n},$$

since $E(b_{22}^2) = 1 - E(a_{12}^2) = 1 - 1/n$, etc.

10. Example of the calculation of the conditional moments. Following Bartlett, we introduce the notation[2]

$$E(\alpha_{11}^2\alpha_{12}^2)E(\beta_{11}^2\beta_{12}^2) = \begin{pmatrix}2\\2\end{pmatrix},$$

(10.1)

$$E(\alpha_{11}^2\alpha_{22}^2)E(\beta_{11}^2\beta_{22}^2) = \begin{pmatrix}2 & \cdot\\ \cdot & 2\end{pmatrix}, \quad \text{etc.}$$

From equation (6.1) it is seen that the conditional moments can be expressed as linear combinations of "arrays" similar to those in (10.1). As we saw in section 6, it is sufficient to calculate the α-moments only.

To illustrate the method let us calculate the α-moment or "half-factor" corresponding to

$$\begin{pmatrix}1 & 1 & \cdot\\ \cdot & 1 & 1\\ 1 & \cdot & 1\end{pmatrix},$$

i.e. $E(\alpha_{11}\alpha_{13}\alpha_{21}\alpha_{22}\alpha_{32}\alpha_{33})$.

The first step is to calculate $\mathfrak{M}m(A)$. Now,

$$C'CA'A = \begin{bmatrix} * & . & c_{21}c_{22} + \cdots . & c_{11}c_{13} + \cdots . & \cdots\\ c_{21}c_{22} + \cdots . & & * & . & c_{32}c_{33} + \cdots . & \cdots\\ c_{11}c_{13} + \cdots . & & c_{32}c_{33} + \cdots . & * & . & \cdots\\ \cdots & \cdots & \cdots & \cdots & \cdots \end{bmatrix} \cdot \begin{bmatrix} 1 & \alpha_1'\alpha_2 & \alpha_1'\alpha_3 & \cdots\\ \alpha_1'\alpha_2 & 1 & \alpha_2'\alpha_3 & \cdots\\ \alpha_1'\alpha_3 & \alpha_2'\alpha_3 & 1 & \cdots\\ \cdots & \cdots & \cdots & \cdots \end{bmatrix}.$$

All remaining terms in $C'C$ may be neglected as they will not contribute to $m(C)$ in the expansion (8.2).

$$\therefore z_1^3 = 48(\alpha_1'\alpha_2)(\alpha_2'\alpha_3)(\alpha_1'\alpha_3)m(C) + \cdots,$$

$$z_1z_2 = 4\{3(\alpha_1'\alpha_2)(\alpha_2'\alpha_3)(\alpha_1'\alpha_3) - (\alpha_1'\alpha_2)^2 - (\alpha_2'\alpha_3)^2 - (\alpha_1'\alpha_3)^2\}m(C) + \cdots,$$

$$z_3 = 2m(C)\begin{vmatrix} 1 & \alpha_1'\alpha_2 & \alpha_1'\alpha_3\\ \alpha_1'\alpha_2 & 1 & \alpha_2'\alpha_3\\ \alpha_1'\alpha_3 & \alpha_2'\alpha_3 & 1\end{vmatrix} + \cdots.$$

Hence, after calculating the expectations of the invariant functions by the method of section 9, we have

$$E(z_1^3) = \frac{48}{n^2}m(C) + \cdots$$

$$E(z_1z_2) = -\frac{12(n-1)}{n^2}m(C) + \cdots$$

$$E(z_3) = \frac{2(n-1)(n-2)}{n^2}m(C) + \cdots.$$

[2] Actually our notation differs slightly from Barlett's. Whereas Bartlett worked in terms of rows vectors, we have worked in terms of column vectors, and hence Bartlett's α_{ij} corresponds to our α_{ji}.

Substituting in (8.2) and equating the coefficients of $m(C)$, we obtain after simplification

$$E(\alpha_{11}\alpha_{13}\alpha_{21}\alpha_{22}\alpha_{32}\alpha_{33}) = \frac{(n-p)(2n-p)}{n^2p(p+2)(p+4)(p-1)(p-2)},$$

which agrees with the value tabulated by Bartlett.

Any other α-moment can be calculated in a similar fashion. In particular, the moments tabulated by Bartlett were checked and the various terms contained in $\mu(2, 1, 1)$ and $\mu(1, 1, 1, 1)$ have been calculated and included in the appendix. Actually, only the α-moments have been tabulated. The complete value for case (a) may be obtained by multiplying the α-moment by a similar value with q replacing p. The complete value for case (b) is obtained by taking the previous value and letting $n \to \infty$ in the second half.

Incidentally, the α-moments may be checked by an independent method. For example, consider the monomial $\alpha_{11}^4\alpha_{12}^2$. If we multiply it by $\alpha_3'\alpha_3$, which is identically unity, then $E[\alpha_{11}^4\alpha_{12}^2(\alpha_3'\alpha_3)] = E[\alpha_{11}^4\alpha_{12}^2]$. But expanding the term on the left-hand side, we get

$$E[\alpha_{11}^4\alpha_{12}^2] = E[\alpha_{11}^4\alpha_{12}^2\alpha_{13}^2] + E[\alpha_{11}^4\alpha_{12}^2\alpha_{23}^2] + E[\alpha_{11}^4\alpha_{12}^2\alpha_{33}^2] + \cdots,$$

and therefore

$$\begin{pmatrix}4\\2\end{pmatrix} = \begin{pmatrix}4\\2\\2\end{pmatrix} + (p-1)\begin{pmatrix}4 & \cdot\\2 & \cdot\\ \cdot & 2\end{pmatrix}.$$

Similarly, by expanding $(\alpha_1'\alpha_2)^2(\alpha_1'\alpha_3)^2$, whose expectation $= 1/n^2$, we have

$$p\begin{pmatrix}4\\2\\2\end{pmatrix} + p(p-1)\begin{pmatrix}2 & 2\\2 & \cdot\\ \cdot & 2\end{pmatrix} + 4p(p-1)\begin{pmatrix}3 & 1\\1 & 1\\2 & \cdot\end{pmatrix} + 2p(p-1)\begin{pmatrix}2 & 2\\1 & 1\\1 & 1\end{pmatrix}$$
$$+ 2p(p-1)(p-2)\begin{pmatrix}2 & 1 & 1\\ \cdot & 1 & 1\\2 & \cdot & \cdot\end{pmatrix} + 4p(p-1)(p-2)\begin{pmatrix}2 & 1 & 1\\1 & 1 & \cdot\\1 & \cdot & 1\end{pmatrix}$$
$$+ p(p-1)(p-2)(p-3)\begin{pmatrix}1 & 1 & 1 & 1\\1 & 1 & \cdot & \cdot\\ \cdot & \cdot & 1 & 1\end{pmatrix} = 1/n^2.$$

REFERENCES

[1] M. S. Bartlett, "The general canonical correlation distribution," *Ann. Math. Stat.*, Vol. 18 (1947), pp. 1–17.

[2] R. A. Fisher, "The general sampling distribution of the multiple correlation coefficient," *Proc. Roy. Soc.*, Vol. A121 (1928), pp. 654–673.

[3] R. A. Fisher, "The sampling distribution of some statistics obtained from non-linear equations," *Ann. Eug.*, Vol. 9 (1939), pp. 238–249.

[4] H. Hotelling, "Relations between two sets of variates," *Biometrika*, Vol. 28 (1936) pp. 321–377.

[5] P. L. Hsu, "On the distribution of roots of certain determinantal equations," *Ann. Eug.*, Vol. 9 (1939), pp. 250–258.

[6] A. T. James, "Normal multivariate analysis and the orthogonal group," *Ann. Math Stat.*, Vol. 25 (1954), pp. 40–75.

[7] A. T. James, "A generating function for averages over the orthogonal group," *Proc. Roy. Soc.*, Vol. A229 (1955), pp. 367–375.

[8] S. N. Roy, "*p*-statistics or some generalisations in analysis of variance appropriate to multivariate problems," *Sankhya*, Vol. 4 (1939), pp. 381–396.

[9] H. Weyl, *The Classical Groups*, Princeton University Press, 2nd ed., 1946.

APPENDIX

$$\begin{pmatrix}4\\2\\2\end{pmatrix} = \frac{3(n+4)(n+6)}{n^2p(p+2)(p+4)(p+6)} \cdot \begin{pmatrix}4 & \cdot\\2 & \cdot\\\cdot & 2\end{pmatrix} = \frac{3(n+4)(np+5n-6)}{n^2p(p+2)(p+4)(p+6)(p-1)}.$$

$$\begin{pmatrix}2 & 2\\2 & \cdot\\2 & \cdot\end{pmatrix} = \frac{(n+4)(np+3n+2p-6)}{n^2p(p+2)(p+4)(p+6)(p-1)} \cdot \begin{pmatrix}3 & 1\\1 & 1\\2 & \cdot\end{pmatrix} = \frac{-3(n-p)(n+4)}{n^2p(p+2)(p+4)(p+6)(p-1)}.$$

$$\begin{pmatrix}4 & \cdot\\1 & 1\\1 & 1\end{pmatrix} = \frac{-3(n-p)(n+4)}{n^2(p+2)(p+4)(p+6)(p-1)}.$$

$$\begin{pmatrix}4 & \cdot\\\cdot & 2\\\cdot & 2\end{pmatrix} = \frac{3\{n^2(p+3)(p+5)+2n(p+1)(p+3)-8(2p+3)\}}{n^2p(p+2)(p+4)(p+6)(p-1)(p+1)}.$$

$$\begin{pmatrix}2 & 2\\2 & \cdot\\\cdot & 2\end{pmatrix} = \frac{n^2(p+3)^2+2n(p+1)(2p+3)+4(p^2-4p-6)}{n^2p(p+2)(p+4)(p+6)(p-1)(p+1)}.$$

$$\begin{pmatrix}3 & 1\\1 & 1\\\cdot & 2\end{pmatrix} = \frac{-3(n-p)(np+3n+2p)}{n^2p(p+2)(p+4)(p+6)(p-1)(p+1)} \cdot \begin{pmatrix}2 & 2\\1 & 1\\1 & 1\end{pmatrix} = \frac{-(n-p)(np-3n+8p+12)}{n^2p(p+2)(p+4)(p+6)(p-1)(p+1)}.$$

$$\begin{pmatrix}4 & \cdot & \cdot\\\cdot & 2 & \cdot\\\cdot & \cdot & 2\end{pmatrix} = \frac{3\{n^2(p^3+8p^2+13p-2)-2n(5p^2+27p+22)+8(5p+6)\}}{n^2p(p+2)(p+4)(p+6)(p-1)(p+1)(p-2)}.$$

$$\begin{pmatrix}2 & 2 & \cdot\\2 & \cdot & \cdot\\\cdot & \cdot & 2\end{pmatrix} = \frac{n^2(p^3+6p^2+3p-6)+2n(p^3-19p-18)-4(3p^2-8p-12)}{n^2p(p+2)(p+4)(p+6)(p-1)(p+1)(p-2)}.$$

$$\begin{pmatrix} \cdot & 2 & 2 \\ 2 & \cdot & \cdot \\ 2 & \cdot & \cdot \end{pmatrix} = \frac{n^2(p+3)(p+5) + 2n(p+1)(p+3) - 8(2p+3)}{n^2p(p+2)(p+4)(p+6)(p-1)(p+1)}.$$

$$\begin{pmatrix} 3 & 1 & \cdot \\ 1 & 1 & \cdot \\ \cdot & \cdot & 2 \end{pmatrix} = \frac{-3(n-p)(np^2 + 5np + 2n - 6p - 4)}{n^2p(p+2)(p+4)(p+6)(p-1)(p+1)(p-2)}.$$

$$\begin{pmatrix} 4 & \cdot & \cdot \\ \cdot & 1 & 1 \\ \cdot & 1 & 1 \end{pmatrix} = \frac{-3(n-p)(np^2 + 7np + 14n - 8p - 16)}{n^2p(p+2)(p+4)(p+6)(p-1)(p+1)(p-2)}.$$

$$\begin{pmatrix} 2 & \cdot & 2 \\ 1 & 1 & \cdot \\ 1 & 1 & \cdot \end{pmatrix} = \frac{-(n-p)(np^2 + 3np + 6n + 2p^2 - 6p - 12)}{n^2p(p+2)(p+4)(p+6)(p-1)(p+1)(p-2)}.$$

$$\begin{pmatrix} 1 & 1 & 2 \\ 1 & 1 & \cdot \\ 2 & \cdot & \cdot \end{pmatrix} = \frac{-(n-p)(np + 3n + 2p)}{n^2p(p+2)(p+4)(p+6)(p-1)(p+1)}.$$

$$\begin{pmatrix} 2 & 1 & 1 \\ 1 & 1 & \cdot \\ 1 & \cdot & 1 \end{pmatrix} = \frac{-(n-p)(3p^2 - 4np + p - 6)}{n^2p(p+2)(p+4)(p+6)(p-1)(p+1)(p-2)}.$$

$$\begin{pmatrix} 2 & 1 & 1 \\ \cdot & 1 & 1 \\ 2 & \cdot & \cdot \end{pmatrix} = \frac{-(n-p)(np^2 + 3np + 6n + 2p^2 - 6p - 12)}{n^2p(p+2)(p+4)(p+6)(p-1)(p+1)(p-2)}.$$

$$\begin{pmatrix} 3 & 1 & \cdot \\ \cdot & 1 & 1 \\ 1 & \cdot & 1 \end{pmatrix} = \frac{3(n-p)(2np + 4n - p^2 - p - 2)}{n^2p(p+2)(p+4)(p+6)(p-1)(p+1)(p-2)}.$$

$$\begin{pmatrix} 2 & 2 & \cdot & \cdot \\ \cdot & \cdot & 2 & \cdot \\ \cdot & \cdot & \cdot & 2 \end{pmatrix} = \frac{n^2(p^3 + 8p^2 + 13p - 2) - 2n(5p^2 + 27p + 22) + 8(5p + 6)}{n^2p(p + 2)(p + 4)(p + 6)(p - 1)(p + 1)(p - 2)}$$

$$\begin{pmatrix} 2 & 2 & \cdot & \cdot \\ \cdot & \cdot & 1 & 1 \\ \cdot & \cdot & 1 & 1 \end{pmatrix} = \frac{-(n - p)(np^2 + 7np + 14n - 8p - 16)}{n^2p(p + 2)(p + 4)(p + 6)(p - 1)(p + 1)(p - 2)}.$$

$$\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & \cdot & \cdot \\ \cdot & \cdot & 1 & 1 \end{pmatrix} = \frac{(n - p)(n - p - 2)}{n^2p(p + 2)(p + 4)(p + 6)(p - 1)(p + 1)}.$$

$$\begin{pmatrix} 2 & 1 & 1 & \cdot \\ \cdot & 1 & 1 & \cdot \\ \cdot & \cdot & \cdot & 2 \end{pmatrix} = \frac{-(n - p)(np^2 + 5np + 2n - 6p - 4)}{n^2p(p + 2)(p + 4)(p + 6)(p - 1)(p + 1)(p - 2)}.$$

$$\begin{pmatrix} 1 & 1 & \cdot & 2 \\ \cdot & 1 & 1 & \cdot \\ 1 & \cdot & 1 & \cdot \end{pmatrix} = \frac{(n - p)(2np + 4n - p^2 - p - 2)}{n^2p(p + 2)(p + 4)(p + 6)(p - 1)(p + 1)(p - 2)}.$$

$$\begin{pmatrix} 2 \\ 2 \\ 2 \\ 2 \end{pmatrix} = \frac{(n + 2)(n + 4)(n + 6)}{n^3p(p + 2)(p + 4)(p + 6)} \cdot \begin{pmatrix} 2 & \cdot \\ 2 & \cdot \\ 2 & \cdot \\ \cdot & 2 \end{pmatrix} = \frac{(n + 2)(n + 4)(np + 5n - 6)}{n^3p(p + 2)(p + 4)(p + 6)(p - 1)}.$$

$$\begin{pmatrix} 1 & 1 \\ 1 & 1 \\ 2 & \cdot \\ 2 & \cdot \end{pmatrix} = \frac{-(n - p)(n + 2)(n + 4)}{n^3p(p + 2)(p + 4)(p + 6)(p - 1)}.$$

$$\begin{pmatrix} 2 & \cdot \\ 2 & \cdot \\ \cdot & 2 \\ \cdot & 2 \end{pmatrix} = \frac{(n + 2)\{n^2(p + 3)(p + 5) + 2n(p + 1)(p + 3) - 8(2p + 3)\}}{n^3(p + 2)(p + 4)(p + 6)(p - 1)(p + 1)}.$$

$$\begin{pmatrix} 2 & \cdot \\ \cdot & 2 \\ 1 & 1 \\ 1 & 1 \end{pmatrix} = \frac{-(n-p)(n+2)(np+3n+2p)}{n^3p(p+2)(p+4)(p+6)(p-1)(p+1)} \cdot \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{pmatrix} \frac{3(n-p)(n+2)(n-p-2)}{n^3p(p+2)(p+4)(p+6)(p-1)(p+1)}.$$

$$\begin{pmatrix} 2 & \cdot & \cdot \\ 2 & \cdot & \cdot \\ \cdot & 2 & \cdot \\ \cdot & \cdot & 2 \end{pmatrix} = \frac{(n+2)\{n^2(p^3+8p^2+13p-2)-2n(5p^2+27p+22)+8(5p+6)\}}{n^3p(p+2)(p+4)(p+6)(p-1)(p+1)(p-2)}.$$

$$\begin{pmatrix} 2 & \cdot & \cdot \\ 1 & 1 & \cdot \\ 1 & 1 & \cdot \\ \cdot & \cdot & 2 \end{pmatrix} = \frac{-(n-p)(n+2)(np^2+5np+2n-6p-4)}{n^3p(p+2)(p+4)(p+6)(p-1)(p+1)(p-2)}.$$

$$\begin{pmatrix} 1 & \cdot & 1 \\ 1 & 1 & \cdot \\ 1 & 1 & \cdot \\ 1 & \cdot & 1 \end{pmatrix} = \frac{(n+2)(n-p)(n-p-2)}{n^3p(p+2)(p+4)(p+6)(p-1)(p+1)}.$$

$$\begin{pmatrix} 1 & 1 & \cdot \\ \cdot & 1 & 1 \\ 1 & \cdot & 1 \\ 2 & \cdot & \cdot \end{pmatrix} = \frac{(n-p)(n+2)(2np+4n-p^2-p-2)}{n^3p(p+2)(p+4)(p+6)(p-1)(p+1)(p-2)}.$$

$$\begin{pmatrix} 2 & \cdot & \cdot \\ \cdot & 1 & 1 \\ \cdot & 1 & 1 \\ 2 & \cdot & \cdot \end{pmatrix} = \frac{-(n-p)(n+2)(np^2+7np+14n-8p-16)}{n^3p(p+2)(p+4)(p+6)(p-1)(p+1)(p-2)}.$$

$$\begin{pmatrix} 2 & \cdot & \cdot & \cdot \\ \cdot & 2 & \cdot & \cdot \\ \cdot & \cdot & 2 & \cdot \\ \cdot & \cdot & \cdot & 2 \end{pmatrix} = \frac{n^3(p^4+7p^3+p^2-35p-6)-12n^2(p^3+6p^2+3p-6)+4n(19p^2+79p+54)-48(5p+6)}{n^3p(p+2)(p+4)(p+6)(p-1)(p+1)(p-2)(p-3)}.$$

$$\begin{pmatrix} 1 & 1 & \cdot & \cdot \\ 1 & 1 & \cdot & \cdot \\ \cdot & \cdot & 1 & 1 \\ \cdot & \cdot & 1 & 1 \end{pmatrix} = \frac{(n-p)\{n^2(p^2+5p+18) - n(p^3+5p^2+18p) + 4(2p^2+3p-6)\}}{n^3p(p+2)(p+4)(p+6)(p-1)(p+1)(p-2)(p-3)}.$$

$$\begin{pmatrix} 1 & 1 & \cdot & \cdot \\ \cdot & 1 & 1 & \cdot \\ 1 & \cdot & 1 & \cdot \\ \cdot & \cdot & \cdot & 2 \end{pmatrix} = \frac{(n-p)\{2n^2(p^2+4p) - n(p^3+4p^2+15p+18) + 6(p^2+p+2)\}}{n^3p(p+2)(p+4)(p+6)(p-1)(p+1)(p-2)(p-3)}.$$

$$\begin{pmatrix} 2 & \cdot & \cdot & \cdot \\ \cdot & 1 & 1 & \cdot \\ \cdot & 1 & 1 & \cdot \\ \cdot & \cdot & \cdot & 2 \end{pmatrix} = \frac{-(n-p)\{n^2(p^3+6p^2+3p-6) - 2n(5p^2+21p+18) + 8(5p+6)\}}{n^3p(p+2)(p+4)(p+6)(p-1)(p+1)(p-2)(p-3)}.$$

$$\begin{pmatrix} 1 & 1 & \cdot & \cdot \\ \cdot & 1 & 1 & \cdot \\ \cdot & \cdot & 1 & 1 \\ 1 & \cdot & \cdot & 1 \end{pmatrix} = \frac{-(n-p)\{n^2(5p+6) - n(5p^2+6p) + (p^3+p^2+2p)\}}{n^3p(p+2)(p+4)(p+6)(p-1)(p+1)(p-2)(p-3)}.$$

SIGNIFICANCE LEVEL AND POWER[1]

By E. L. Lehmann

University of California, Berkeley

1. Summary and introduction. Significance testing, as described in most textbooks, consists in fixing a standard significance level α such as .01 or .05 and rejecting the hypothesis $\theta = \theta_0$ if a suitable statistic Y exceeds C where $P_{\theta_0}\{Y > C\} = \alpha$. Such a procedure controls the probability of false rejection (error of the first kind) at the desired level α but leaves the power of the test and hence the probability of an error of the second kind to the mercy of the experiment. It seems more natural when deciding on a significance level (and this suggestion is certainly not new) to take into account also what power can be achieved with the given experiment. In Section 3 a specific suggestion will be made as to how to balance α against the power β obtainable against the alternatives of interest.

The adoption of this or some similar rule for choosing a significance level has important consequences for the theory of testing composite hypotheses, where nuisance parameters are present. Since the quantity α is then potentially a function of the nuisance parameter ϑ, the classical rule of a fixed significance level leads to the condition that the tests be *exact* or *similar*, that is, that $\alpha(\vartheta)$ equal the preassigned value α for all ϑ. On the other hand, the power β that can be attained against any alternative $\theta = \theta_1$ frequently depends on ϑ. The requirement that $\alpha(\vartheta)$ and $\beta(\vartheta)$ be in a certain balance thus leads to tests which are not similar and hence do not agree with the standard solutions.

To obtain a suitable setting for this discussion, we consider first a minimal complete class of tests for testing the hypothesis $H:\theta \leqq \theta_0$ in a multiparameter exponential family (Section 2). The proposed α, β-relation is discussed in Section 3, and in Section 4 is applied to the exponential family. Section 5 gives some illustrations of the theory.

2. A complete class theorem. Many standard testing problems concern an exponential family of distributions, which has probability densities of the form

$$p_{\theta,\vartheta}(x) = C(\theta, \vartheta) \exp\left[\theta U(x) + \sum_{i=1}^{r} \vartheta_i T_i(x)\right] h(x) \tag{1}$$

with respect to a σ-finite measure μ, where θ, U, the ϑ_i and T_i are real-valued and where $\vartheta = (\vartheta_1, \cdots, \vartheta_r)$. In this family, the statistics U and $T = (T_1, \cdots, T_r)$ constitute a set of sufficient statistics for (θ, ϑ).

The problem of testing the hypothesis $H:\theta \leqq \theta_0$ against the one-sided al-

Received January 17, 1958; revised July 22, 1958.

[1] This paper was prepared with the partial support of the Office of Naval Research. This paper in whole or in part may be reproduced for any purpose of the United States Government.

ternatives $\theta > \theta_0$ has been treated by many authors (usually in the formulation $\theta = \theta_0$ against $\theta > \theta_0$). The solution of this testing problem according to the Neyman-Pearson theory is the uniformly most powerful unbiased test; this depends only on U and T and is given by the critical function[2]

$$\phi(u, t) = \begin{cases} 1 \text{ if } u > C(t), \\ \gamma(t) \text{ if } u = C(t), \\ 0 \text{ if } u < C(t), \end{cases} \tag{2}$$

where the functions C and γ are determined by the conditions $E_{\theta_0}[\phi(U, T) \mid T = t] = \alpha$ and $E_{\theta_0}[U\phi(U, T) \mid T = t] = \alpha\, E_{\theta_0}[U \mid T = t]$ for all t. The condition of unbiasedness

$$E_{\theta,\vartheta}\phi(U, T) \lesseqqgtr \alpha \qquad \text{as} \qquad \theta \lesseqqgtr \theta_0,$$

and that of similarity

$$E_{\theta_0,\vartheta}\phi(U, T) = \alpha \qquad \text{for all } \vartheta$$

which it implies and which by itself is sufficient to justify the test, are not inherent in the problem but are imposed, at least in part, to facilitate the solution. Before proposing an alternative approach, it is interesting to see how far the problem can be reduced without the introduction of extraneous principles. This can be done by viewing it within the framework of decision theory.

Let d_0 and d_1 denote the decisions of accepting and rejecting the hypothesis H, and denote by $L_i(\theta, \vartheta)$ the loss resulting from decision d_i when (θ, ϑ) are the true parameter values. Then for fixed ϑ, the function $L_0(\theta, \vartheta)$ typically will be zero for $\theta \leqq \theta_0$ and increasing for $\theta \geqq \theta_0$, while $L_1(\theta, \vartheta)$ will be decreasing for $\theta \leqq \theta_0$ and zero for $\theta \geqq \theta_0$. In particular, the difference then satisfies

$$L_1(\theta, \vartheta) - L_0(\theta, \vartheta) \gtreqqless 0 \qquad \text{as} \qquad \theta \lesseqqgtr \theta_0. \tag{3}$$

The risk function of a test ϕ, which is the expected loss resulting from its use considered as a function of the parameters, is

$$\begin{aligned} R_\varphi(\theta, \vartheta) = \int \{&\varphi(U(x), T(x))L_1(\theta, \vartheta) \\ &+ [1 - \varphi(U(x), T(x))]L_0(\theta, \vartheta)\}\, p_{\theta,\vartheta}(x)\, d\mu(x). \end{aligned} \tag{4}$$

Let $\mathcal{C}$ be the class of all tests satisfying (2) for some functions C and γ. For all loss functions satisfying (3) it was shown by Truax [13] that $\mathcal{C}$ is essentially complete; that is, given any φ there exists $\varphi' \varepsilon\, \mathcal{C}$ such that

$$R_{\varphi'}(\theta, \vartheta) \leqq R_\varphi(\theta, \vartheta) \qquad \text{for all } (\theta, \vartheta). \tag{5}$$

We shall now prove that among essentially complete classes, $\mathcal{C}$ is minimal in the sense that if (5) holds for two tests φ, φ' in $\mathcal{C}$, then $\varphi = \varphi'$ a.e. μ.*

---

[2] See for example [7].

* Recently I learned that this result has been obtained also by D. L. Burkholder. His results are sketched in Abstract 18, *Ann. Math. Stat.*, Vol. 29 (1958), p. 616.

Let φ and φ' belong to $\mathcal{C}$ and let

$$\alpha(\vartheta) = E_{\theta_0,\vartheta}\,\varphi(U, T), \qquad \alpha'(\vartheta) = E_{\theta_0,\vartheta}\,\varphi'(U, T). \tag{6}$$

(i) If the functions α and α' do not agree for all ϑ, suppose without loss of generality that there exists ϑ_0 such that $\alpha(\vartheta_0) < \alpha'(\vartheta_0)$. Since for $\vartheta = \vartheta_0$, the expected values of φ and φ' are continuous functions of θ, there exist $\theta_1 < \theta_0 < \theta_2$ such that

$$E_{\theta,\vartheta_0}\varphi(U, T) < E_{\theta,\vartheta_0})\varphi'(U, T) \qquad \text{for } \theta = \theta_1 \text{ and } \theta = \theta_2 . \tag{7}$$

Then $R_\phi(\theta_1, \vartheta_0) < R_{\phi'}(\theta_1, \vartheta_0)$ and $R_\phi(\theta_2, \vartheta_0) > R_{\phi'}(\theta_2, \vartheta_0)$, and hence neither of the procedures ϕ and ϕ' is uniformly better than the other. (ii) Suppose on the other hand that $\alpha(\vartheta) \equiv \alpha'(\vartheta)$. The standard proof showing a similar test satisfying (2) to be uniformly most powerful similar also shows that a test ϕ_0 satisfying (2) and

$$E_{\theta_0,\vartheta}\,\phi_0(U, T) = \alpha(\vartheta) \qquad \text{for all } \vartheta \tag{8}$$

is uniformly most powerful among all tests satisfying (8). The tests ϕ and ϕ' are therefore both uniformly most powerful within this class and hence

$$E_{\theta,\vartheta}\,\phi(U, T) = E_{\theta,\vartheta}\,\phi'(U, T) \qquad \text{for all } \theta > \theta_0 \text{ and all } \vartheta.$$

Since the family of distributions of the sufficient statistics (U, T) is complete, it follows that $\phi(u, t) = \phi'(u, t)$ a.e., as was to be proved.

3. Significance level and power. It follows from the result of the preceding section that the class $\mathcal{C}$ of tests (2) represents the maximum reduction that can be achieved by comparing only tests of which one has a uniformly better risk function than the other. The selection of a specific test from $\mathcal{C}$, involves two difficulties. It requires the adoption of some principle (Bayes, minimax, etc.) leading to a definite choice;[3] in addition, it requires knowledge of the loss functions L_0 and L_1. An alternative approach, utilizing the fortunate circumstance that the complete class is independent of the actual loss functions (subject only to their satisfying (3)), consists in making the choice by some simple rule of thumb, which does not require (the usually unavailable) knowledge of these losses.

Consider the simplest case of the family (1) with $r = 0$, which involves no nuisance parameters. The family of tests (2) is then a one-parameter family, one test corresponding to each value of

$$\alpha_0 = E_{\theta_0}\phi(X), \qquad 0 \leqq \alpha_0 \leqq 1.$$

A simple method of choice consists in specifying a value of α_0 and selecting the test corresponding to this value. This need not be a purely formal or arbitrary

[3] Particular proposals of this kind that have been made in the literature include those of Jeffreys [5] involving considerations of *a priori* probabilities, and of Lindley [8] based on his concept of unlikelihood.

procedure since α_0 as the maximum probability of false rejection is of course an important quantity in its own right.

Nevertheless, as was pointed out in Section 1, the above rule appears to neglect too many aspects of the problem. In particular, suppose that the alternatives of primary interest, for which it is important to reject the hypothesis, are those satisfying $\theta \geqq \theta_1$ ($\theta_0 < \theta_1$). Since the power function of any test (2) is increasing in θ, the probability β_1 of rejection when $\theta = \theta_1$ is the minimum power against these alternatives. It seems then reasonable that the choice of test should involve at least β_1 in addition to α_0.

The quantities α_0 and $\alpha_1 = 1 - \beta_1$ are the error probabilities associated with the problem of testing the simple hypothesis $\theta = \theta_0$ against the simple alternative $\theta = \theta_1$. The attainable pairs (α_0, α_1) form a convex set, the lower boundary of which corresponds to the admissible tests (2). This lower boundary is a convex curve S connecting the points (0, 1) and (1, 0), and what is needed is a reasonable way of selecting a point on each such curve. One possible approach to this question is in terms of indifference curves. Suppose that a system of curves could be specified in the (α_0, α_1)-plane such that any two points lying on the same curve are equally desirable, with the curves closer to the origin being more desirable than those further away. The optimum test would then be given by that point of S lying on the indifference curve closest to the origin (Fig. 1).

It seems likely that even this approach is too complex for most applications. To obtain an even simpler formulation, consider once more the rule of fixing the significance level without regard to power. If the level is α, this means restricting attention to the points (α_0, α_1) lying on the vertical line segment $L: \alpha_0 = \alpha$, $0 \leqq \alpha_1 \leqq 1 - \alpha$. The test then corresponds to the point (α_0, α_1), which is the

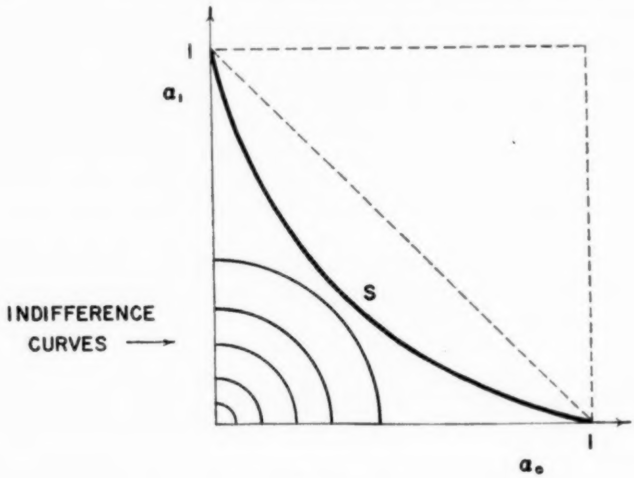


FIG. 1

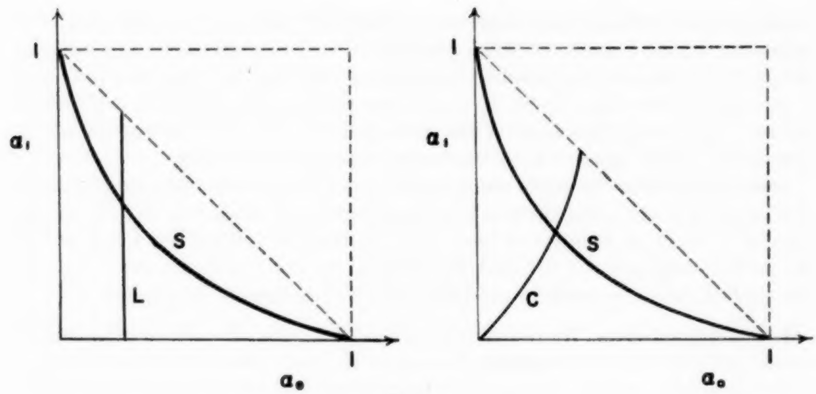


FIG. 2

intersection of S and L. This procedure is commonly justified on the grounds that the error of the first kind is of a higher order of importance, and should therefore be controlled at the prescribed level. However, if the curve S is sufficiently close to the α_0- and α_1-axis, as will always be the case if the sample size is sufficiently large, then α_1 is much smaller than α_0, which is inconsistent with the assumed relative importance of the two errors.

A more reasonable solution is obtained if one replaces the vertical line segment L by a curve $C: \alpha_1 = f(\alpha_0)$ where f is a continuous strictly increasing function with $f(0) = 0$. A particularly simple choice for f is a linear function

$$\alpha_1 = k\alpha_0 . \tag{9}$$

Since $\alpha_0 \leqq 1 - \alpha_1$ for all admissible tests, one has $\alpha_0 \leqq 1/(k+1)$ so that $1/(k+1)$ is an upper bound for α_0. As an example, consider (9) with $k = 9$. If $\beta_1 = 1 - \alpha_1$ denotes the power of a test against the alternative θ_1, some typical pairs of values of (α_0, β_1) are

α_0	.1	.05	.04	.03	.02	.01	.005
β_1	.1	.55	.64	.73	.82	.91	.955

with .1 being an upper bound for α_0.

One would of course hope to avoid cases such as $\alpha_0 = .1$, $\beta_1 = .1$ or even $\alpha_0 = .05$, $\beta_1 = .55$. When no nuisance parameters are present, this can be achieved by taking a sample of sufficient size. In the composite case, on the other hand, it can frequently not be achieved by samples of fixed size no matter how large, but only by resorting to sequential experimentation.

To avoid misunderstandings, it should be emphasized that (9) is not being proposed as a logically convincing rule, nor as one fitting all occasions. Actually, it seems clear that no rule satisfying these requirements exists, except the Bayes

solution when sufficient knowledge concerning losses and *a priori* probabilities is available. In the absence of this knowledge it may be convenient to employ a simple rule of thumb. Such a rule is in fact being used in much of present practice: It consists in choosing α to be .05 or .01 depending on the seriousness attached to the committing of an error of the first kind. To this, (9) is suggested as an alternative which appears to be more reasonable in many cases.

It so happens that (9) is the minimax solution if the loss for rejecting $H: \theta \leqq \theta_0$ is a_0 when H is true, and the loss is a_1 for accepting H when $\theta \geqq \theta_1$, where the constant k of (9) is then given by $k = a_0/a_1$. However, this is not the basis for the present suggestion of (9), and the minimax property does not carry over to the application to be made in the next section to composite hypotheses.

4. Conditional tests. We return now to the composite case of the exponential family (1) with $r > 0$. The minimal complete class $\mathcal{C}$ is then more complex than in the preceding section, its members being characterized by the function $\alpha(\vartheta)$ instead of the single number α_0. Given any function $\alpha(\vartheta)$, which is the expectation of some critical function ϕ, there exists a unique member of $\mathcal{C}$ whose expectation function for $\theta = \theta_0$ is also $\alpha(\vartheta)$. This uniformly minimizes the risk (and maximizes the power) among all critical functions having this expectation.

If the alternatives of interest are as before those satisfying $\theta' \geqq \theta_1$, let $\beta(\vartheta)$ denote the power function of a test against the alternative (θ_1, ϑ). The proposal made in the preceding section suggests selecting that member of $\mathcal{C}$ which satisfies

$$1 - \beta(\vartheta) = k\alpha(\vartheta) \qquad \text{for all } \vartheta. \tag{10}$$

However, this relationship depends on the particular parametrization chosen, and we shall not discuss it here. Instead an alternative approach will be proposed in which this difficulty does not arise.

Consider once more the case of the similar test with $\alpha(\vartheta) \equiv \alpha$. Since T is a complete sufficient statistic for ϑ when $\theta = \theta_0$, the functions C and γ of (2) are determined by the requirement that the conditional probability of rejection

$$\alpha^*(t) = P_{\theta_0}\{U > C(t) \mid t\} + \gamma(t)P_{\theta_0}\{U = C(t) \mid t\}$$

be equal to α for all t.[4] However, the conditional power $\beta^*(t) = P_{\theta_1}\{\text{rejecting } H \mid t\}$ of the test against the alternative $\theta = \theta_1$, typically depends on t. The question then arises: Suppose that $\beta^*(t)$ is quite small for the observed t, or quite high; is this value not more relevant to the case in hand than the average value $\beta(\vartheta)$?

Without entering into the difficulties raised by this question, there is an alternative and simpler justification for considering $\beta^*(t)$. The actual power β against the alternative $\theta = \theta_1$ generally depends on the nuisance parameter ϑ and is therefore unknown. It can however be estimated from the observations,

[4] This method of constructing exact tests was originated by Bartlett [1] and Neyman [9]. That in the present case it provides the totality of such tests has been noted by many authors. For a recent discussion and references see [7].

and $\beta^*(T)$ is the unbiased estimate with (uniformly) minimum variance. That it is unbiased is clear since $\beta(\vartheta) = E_{\theta_1,\vartheta}\beta^*(T)$. The minimum variance property is an immediate consequence of the completeness of the sufficient statistic T for (θ_1, ϑ) and of Theorem 5.1 of [7].

Analogous remarks apply in the more general case, in which the tests are not required to be exact. If the relevant frame of reference is obtained by considering t as fixed, the error probabilities of interest are the conditional probabilities $\alpha_0^*(t) = P_{\theta_0}$ (rejecting $H \mid t$) and $\alpha_1^*(t) = P_{\theta_1}$ (accepting $H \mid t$), and the quantities $C(t)$ and $\gamma(t)$ can therefore be determined from the relation

$$\alpha_1^*(t) = k\alpha_0^*(t). \tag{11}$$

The resulting test will of course not be similar. However, since $\alpha_0^*(t) \leqq 1/(k+1)$ for all t, the quantity $1/(k+1)$ is an upper bound also for the average probability $\alpha_0(\vartheta)$ of an error of the first kind.

The above discussion applies only to problems in which the parameter of interest is one of the "natural" parameters of the exponential distribution (1). As was pointed out in [7], any parameter of the form $\theta + \sum a_i\vartheta_i$ is natural for a suitable definition of U, the T's and ϑ's. When the parameter of interest is not of this form, related methods may be applicable as is indicated by the following example.

If $X_1, \cdots, X_n$ are a sample from a normal distribution $N(\xi, \sigma^2)$, neither the parameter ξ nor ξ/σ are of this form. The problem of testing $\xi/\sigma \leqq \delta_0$ against $\xi/\sigma \geqq \delta_1$ can be reduced by invariance considerations to the statistic $\bar{X}/[\sum(X_i - \bar{X})^2]^{\frac{1}{2}}$, the distribution of which depends on the single parameter $\delta = \xi/\sigma$. If $\alpha_i = P_{\delta_i}\{\bar{X} > C[\sum(X_i - \bar{X})^2]^{\frac{1}{2}}\}$, the quantity C can be determined so that $\alpha_1 = k\alpha_0$. The problem of testing $\xi \leqq \xi_0$ against $\xi \geqq \xi_1$ appears to be more difficult; a possible approach may be that of [4], Section 3.

5. Examples. We shall now briefly indicate some examples in which the natural parameter θ is the relevant one so that the method of the preceding sections is applicable. Of these, Examples 1, 2, 3 have been treated by the same method (but from a different point of view) by Tocher [12], and Examples 2, 3 by Sverdrup [11].

EXAMPLE 1. Let X, Y be independent Poisson variables with $E(X) = \lambda$, $E(Y) = \mu$, and consider the problem of testing $\mu/\lambda \leqq a_0$ against $\mu/\lambda \geqq a_1$. The joint distribution of X, Y forms an exponential family with $T = X + Y$, $U = Y$, $\theta = \log(\mu/\lambda)$ and $\vartheta = \log \lambda$. The conditional distribution of Y given $X + Y = t$ is a binomial distribution corresponding to the success probability $p = \mu/(\lambda + \mu)$ and number of trials equal to t. In terms of p, the hypothesis and class of alternatives becomes $p \leqq a_0/(a_0 + 1)$ and $p \geqq a_1/(a_1 + 1)$ so that the test satisfying (2) and (11) can be determined from a table of the binomial distribution.

EXAMPLE 2. If X, Y are independent variables with binomial distributions $b(p_1, m)$ and $b(p_2, n)$, their joint distribution has the exponential form (1) with

$T = X + Y$, $U = Y$, $\theta = \log (p_2/q_2 \div p_1/q_1)$ and $\vartheta = \log (p_1/q_1)$. The method is therefore applicable to the problem of testing $p_2/q_2 \leqq a_0(p_1/q_1)$, and in particular $p_2 \leqq p_1$ by letting $a_0 = 1$, against the alternatives $p_2/q_2 \geqq a_1(p_1/q_1)$. Putting $\rho = (p_2/q_2) \div (p_1/q_1)$, the conditional distribution of Y given t is

$$(12) \qquad P_\rho\{Y = y \mid X + Y = t\} = C_t(\rho) \binom{m}{t-y} \binom{n}{y} \rho^y, \qquad y = 0, 1, \cdots, t,$$

which for $\rho = 1$ reduces to the hypergeometric distribution.

EXAMPLE 3. In a 2×2 table representing the results of classifying a sample of size s according to two characteristics A and B, the joint distribution of the numbers X, Y, Y' in the

	A	$\bar{A}$	
B	X	X'	M
$\bar{B}$	Y	Y'	N
	T	T'	S

categories AB, $A\bar{B}$ and $\bar{A}\bar{B}$ constitute an exponential family with $U = Y$, $T_1 = X + Y$, $T_2 = Y + Y'$ and $\theta = \log (p_{A\bar{B}}p_{\bar{A}B}/p_{AB}p_{\bar{A}\bar{B}})$. Putting $\Delta = (p_{A\bar{B}}p_{\bar{A}B}/p_{AB}p_{\bar{A}\bar{B}})$ one finds

$$p_{AB} = p_A p_B + \frac{1-\Delta}{\Delta} p_{\bar{A}B}p_{A\bar{B}}; \qquad p_{\bar{A}\bar{B}} = p_{\bar{A}} p_{\bar{B}} + \frac{1-\Delta}{\Delta} p_{\bar{A}B}p_{A\bar{B}}$$

$$p_{A\bar{B}} = p_A p_{\bar{B}} - \frac{1-\Delta}{\Delta} p_{\bar{A}B}p_{A\bar{B}}; \qquad p_{\bar{A}B} = p_{\bar{A}} p_B - \frac{1-\Delta}{\Delta} p_{\bar{A}B}p_{A\bar{B}}$$

where p_{AB} denotes the probability of having the characteristics A and B, $p_A = p_{AB} + p_{A\bar{B}}$ the probability of having the characteristic A, etc. The quantity Δ is therefore a measure of the degree of dependence,[5] $\Delta = 1$ corresponding to independence, $\Delta < 1$ to negative and $\Delta > 1$ to positive dependence. The method of the preceding section is applicable to testing $\Delta \leqq 1$ or more generally $\Delta \leqq \Delta_0$ against the alternatives $\Delta \geqq \Delta_1$. The conditional distribution of Y given $X + Y = t$, $Y + Y' = n$ is given by (12) with Δ in place of ρ.

EXAMPLE 4. Consider a number of paired comparisons (U_k, V_k) where only the sign of the differences $W_k = V_k - U_k$ are observed for each pair $k = 1, \cdots, n$. If the probability of a positive, negative and zero observation are p_+, p_- and p_0 in each case and if the comparisons are independent, the joint distribution of the numbers X, Y and Z of positive, negative and zero cases is the multinomial distribution

$$\frac{n!}{x!\,y!\,z!} p_+^x p_-^y p_0^z.$$

---

[5] Δ is equivalent to Yule's measure of association, which is $Q = (1 - \Delta)/(1 + \Delta)$. For a discussion of this and related measures, see [2].

This is an exponential family with $U = Y$, $T = Z$, $\theta = \log(p_+/p_-)$ and $\vartheta = \log(p_0/p_-)$. The test of $p_+ \leqq p_-$ (or $p_+ \leqq a_0 p_-$) against $p_+ \geqq a_1 p_-$ is therefore performed conditionally given $Z = t$. Since the conditional distribution of Y given $Z = t$ is the binomial distribution $b(p_+/(p_+ + p_-), n - t)$, the constants $C(t)$ and $\gamma(t)$ for which the test satisfies (2) and (11) can be obtained from the binomial tables.[6]

Example 5. Let $Y_1, \cdots, Y_N$ be independently distributed according to the binomial distributions $b(p_i, n_i)$ $i = 1, \cdots, N$ where

$$p_i = 1/[1 + e^{-(\alpha+\beta x_i)}]$$

This is the model frequently assumed in bioassay, where x_i denotes the dose or some function of the dose such as its logarithm, of a drug given to n_i experimental subjects and where Y_i is the number among these subjects which respond to the drug at level x_i. Here the x_i are known, and α and β are unknown parameters. The joint distribution of the Y's is

$$e^{\alpha \Sigma y_i + \beta \Sigma x_i y_i} \prod_{i=1}^{N} \binom{n_i}{y_i} \left[\frac{e^{-(\alpha+\beta x_i)}}{1 + e^{-(\alpha+\beta x_i)}} \right]^{n_i}, \tag{13}$$

which is an exponential family with the parameters α, β and sufficient statistics $\sum Y_i$, $\sum x_i Y_i$. The method is therefore applicable to testing $\alpha \leqq \alpha_0$ against $\alpha \geqq \alpha_1$ or $\beta \leqq \beta_0$ against $\beta \geqq \beta_1$. It is interesting to note that for the particular case $x_i = ic$ and $H: \beta \leqq 0$, the conditional test given $Y = t$ is a form of the Wilcoxon test in a setting similar to that discussed by Haldane and Smith [3].

As a last example we mention without going into details the comparison of two distributions of type (13). If the parameters in these are α, β and α', β' the differences $\alpha' - \alpha$ and $\beta' - \beta$ are natural parameters of the resulting exponential families, and can therefore be tested by the method discussed here.

REFERENCES

[1] M. S. Bartlett, "Properties of sufficiency and statistical tests," *Proc. Roy. Soc. London*, Vol. 160 (1937), pp. 268-281.

[2] Leo A. Goodman and William H. Kruskal, "Measures of association for mass classifications," *J. Amer. Stat. Assn.*, Vol. 49 (1954), pp. 732-764.

[3] J. B. S. Haldane and Cedric A. B. Smith, "A simple exact test for birth order effect," *Ann. Eugenics*, Vol. 14 (1948), pp. 117-124.

[4] J. L. Hodges, Jr. and E. L. Lehmann, "Testing the approximate validity of statistical hypotheses," *J. Roy. Stat. Soc., Ser. B.*, Vol. 16 (1954), pp. 261-268.

[5] H. Jeffreys, "Theory of probability (2d ed.), Oxford: Clarendon Press, 1948.

[6] Samuel Karlin and Herman Rubin, "The theory of decision procedures for distribu-

[6] This problem has been considered previously in [10]. The statement made there that the test satisfying (2) and $\alpha^*(t) \equiv \alpha$ is uniformly most powerful is too strong. The test is however uniformly most powerful among all similar (and hence all unbiased) tests. An analogous remark applies to Example 2, which is among those considered in [12]. As was proved there, the test which is conditionally unbiased at fixed level α for each t, is uniformly unbiased; it is however not uniformly most powerful without this restriction as is claimed in [12] for all the cases treated there.

tions with monotone likelihood ratio," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 272–299.

[7] E. L. Lehmann and Henry Scheffé, "Completeness, similar regions and unbiased estimation," *Sankhya*, Vol. 10 (1950), pp. 305–340 and Vol. 15 (1955), pp. 219–236.

[8] D. V. Lindley, "Statistical inference," *J. Roy. Stat. Soc., Ser. B.*, Vol. 15 (1953), pp. 30–76.

[9] J. Neyman, "Outline of a theory of statistical estimation based on the classical theory of probability," *Philos. Trans. Roy. Soc. London*, Vol. 236 (1937), pp. 333–380.

[10] Joseph Putter, "The treatment of ties in some nonparametric tests," *Ann. Math. Stat.*, Vol. 26 (1955), pp. 368–386.

[11] Erling Sverdrup, "Similarity, unbiasedness, minimaxibility, and admissibility of statistical test procedures," *Skand. Aktuarietids.*, Vol. 36 (1953), pp. 64–86.

[12] K. D. Tocher, "Extension of Neyman-Pearson theory of tests to discontinuous variates," *Biometrika*, Vol. 37 (1950), pp. 130–144.

[13] Donald R. Truax, "Multidecision problems for the multivariate exponential family," Stanford Technical Report No. 32 (1955) (to be published).

STEP-DOWN PROCEDURE IN MULTIVARIATE ANALYSIS[1]

By J. Roy

University of North Carolina

1. Introduction and summary. Test criteria for (i) multivariate analysis of variance, (ii) comparison of variance-covariance matrices, and (iii) multiple independence of groups of variates when the parent population is multivariate normal are usually derived either from the likelihood-ratio principle [6] or from the "union-intersection" principle [2]. An alternative procedure, called the "step-down" procedure, has been recently used by Roy and Bargmann [5] in devising a test for problem (iii). In this paper the step-down procedure is applied to problems (i) and (ii) in deriving new tests of significance and simultaneous confidence-bounds on a number of "deviation-parameters."

The essential point of the step-down procedure in multivariate analysis is that the variates are supposed to be arranged in descending order of importance. The hypothesis concerning the multivariate distribution is then decomposed into a number of hypotheses—the first hypothesis concerning the marginal univariate distribution of the first variate, the second hypothesis concerning the conditional univariate distribution of the second variate given the first variate, the third hypothesis concerning the conditional univariate distribution of the third variate given the first two variates, and so on. For each of these component hypotheses concerning univariate distributions, well known test procedures with good properties are usually available, and these are made use of in testing the compound hypothesis on the multivariate distribution. The compound hypothesis is accepted if and only if each of the univariate hypotheses are accepted. It so turns out that the component univariate tests are independent, if the compound hypothesis is true. It is therefore possible to determine the level of significance of the compound test in terms of the levels of significance of the component univariate tests and to derive simultaneous confidence-bounds on certain meaningful parametric functions on the lines of [3] and [4].

The step-down procedure obviously is not invariant under a permutation of the variates and should be used only when the variates can be arranged on a priori grounds. Some advantages of the step-down procedure are (i) the procedure uses widely known statistics like the variance-ratio, (ii) the test is carried out in successive stages and if significance is established at a certain stage, one can stop at that stage and no further computations are needed, and (iii) it leads to simultaneous confidence-bounds on certain meaningful parametric functions.

1.1 *Notations.* The operator $\mathcal{E}$ applied to a matrix of random variables is used

---

Received January 17, 1958; revised May 5, 1958.

[1] This research was supported by the United States Air Force through the Air Force Office of Scientific Research of the Air Research and Development Command. Reproduction in whole or in part is permitted for any purpose of the United States Government.

to generate the matrix of expected values of the corresponding random variables. The form of a matrix is denoted by a subscript; thus $A_{n \times m}$ indicates that the matrix A has n rows and m columns. The maximum latent root of a square matrix B is denoted by $\lambda_{\max}(B)$. Given a vector $a = (a_1, a_2, \cdots, a_t)'$ and a subset T of the natural numbers 1, 2, $\cdots$, t, say $T = (j_1, j_2, \cdots, j_u)$ where $j_1 < j_2 < \cdots j_u$, the notation $T[a]$ will be used to denote the positive quantity:

$$T[a] = + \{a_{j_1}^2 + a_{j_2}^2 + \cdots + a_{j_t}^2\}^{1/2}.$$

$T[a]$ will be called the T-norm of a. Similarly, given a matrix $B_{t \times t}$, we shall write $B_{(T)}$ for the $u \times u$ submatrix formed by taking the j_1th, j_2th, $\cdots$, j_uth rows and columns of B. We shall call $B_{(T)}$ the T-submatrix of B.

2. Step-down procedure in multivariate analysis of variance.

2.1 *General linear hypothesis in univariate analysis.* Let the elements of $y_{n \times 1}$ be one-dimensional random variables distributed independently and normally with the same variance σ^2 and expectations given by

$$\mathcal{E}y = A\theta + X\beta \tag{1}$$

where elements of $\theta_{m \times 1}$ and $\beta_{q \times 1}$ are unknown parameters; $A_{n \times m}$ and $X_{n \times q}$ are matrices of known constants with rank $(A) = r$ and rank $(A:X) = r + q$, with $n > (r + q)$.

A set of t linearly independent linear functions $\phi_{t \times 1} = B_{t \times m}\theta$, where B is a given matrix of rank t, is said to be estimable if for each element of ϕ there exists an unbiassed estimate linear in y, for all values of θ and β. If ϕ is estimable, there exists an estimator $\hat{\phi}_{t \times 1}$ of ϕ, the elements of which are linear in y and minimum variance unbiassed estimators of the corresponding elements in ϕ. Denote the variance-covariance matrix of $\hat{\phi}$ by $C \cdot \sigma^2$, where $C_{t \times t}$ is a positive-definite matrix. Let $s^2/(n - q - r)$ denote the usual error mean square with $(n - q - r)$ degrees of freedom giving an unbiassed estimator of σ^2. Then it is well known that the statistics $u = (\hat{\phi} - \phi)'C^{-1}(\hat{\phi} - \phi)/\sigma^2$ and $v = s^2/\sigma^2$ are distributed independently as chi-squares with t and $(n - q - r)$ degrees of freedom respectively, so that

$$F \equiv \frac{(\hat{\phi} - \phi)'C^{-1}(\hat{\phi} - \phi)/t}{s^2/(n - q - r)} \tag{2}$$

is distributed as a variance-ratio with t and $(n - q - r)$ degrees of freedom.

Let α be a preassigned constant, $0 < \alpha < 1$, and f the upper 100α per cent point of the variance-ratio distribution with t and $(n - q - r)$ degrees of freedom. Setting $\mathfrak{L}^2 = tf/(n - q - r)$ we then have

$$(\hat{\phi} - \phi)'C^{-1}(\hat{\phi} - \phi) \leqq \mathfrak{L}^2 s^2 \tag{3}$$

with probability $(1 - \alpha)$.

Now, the left-hand side of (3) is a positive definite quadratic form in $(\hat{\phi} - \phi)$ and consequently, we have

$$(\hat{\phi} - \phi)'C^{-1}(\hat{\phi} - \phi) \geqq (\hat{\phi} - \phi)'(\hat{\phi} - \phi)/\lambda_{\max}(C). \tag{4}$$

We thus have

$$(\hat{\phi} - \phi)'(\hat{\phi} - \phi) \leqq f^2 s^2 \lambda_{\max}(C) \tag{5}$$

with probability not less than $(1 - \alpha)$.

Now, let T be any subset of the natural numbers $1, 2, \cdots, t$ and consider the T-norms $T[\phi]$ of ϕ and $T[\hat{\phi}]$ of $\hat{\phi}$. Then (3) implies that

$$T[\hat{\phi}] - fs\,\lambda_{\max}^{1/2}(C_{(T)}) \leqq T[\phi] \leqq T[\hat{\phi}] + fs\,\lambda_{\max}^{1/2}(C_{(T)}) \tag{6}$$

for all subsets T of $(1, 2, \cdots, t)$, where C_T is the T-submatrix of C. The statement (6) thus provides simultaneous confidence-bounds on the parameters $T[\phi]$ for all T with probability not less than $(1 - \alpha)$. We note that there are in all $(2^t - 1)$ parameters of the type $T[\phi]$ and these in a sense measure the deviations from the hypothesis $\mathcal{H}_0$ that $\phi = 0$. The analysis of variance test for $\mathcal{H}_0$ at level of significance α, of course, is given by the rule

$$\text{accept } \mathcal{H}_0 \text{ if } \frac{\hat{\phi}'C^{-1}\hat{\phi}/t}{s^2/(n - q - r)} \leqq f; \tag{7}$$
$$\text{otherwise reject } \mathcal{H}_0.$$

However, simultaneous confidence-bounds of the type (6) are more interesting than the test (7) itself, because the direction of departure from the null hypothesis is indicated.

2.2 *Customary tests in multivariate analysis of variance.* We have a matrix $Y_{n \times p}$ of random variables, such that the rows are distributed independently, each row having a p-variate normal distribution with the same variance-covariance matrix $\Sigma_{p \times p}$ which is positive-definite. The expected values are given by

$$\mathcal{E}Y = A\Theta, \tag{8}$$

where $A_{n \times m}$ is a matrix of known constants of rank r, $r \leqq (n - p)$, and $\Theta_{m \times p}$ is a matrix of unknown parameters. As before, a set of linear parametric functions $\Phi_{t \times p} = B_{t \times m}\Theta$ is said to be estimable if, for all Θ, there exist unbiassed estimates of Φ linear in Y. If Φ is estimable, customary tests for the hypothesis

$$\mathcal{H}_0\colon \Phi = 0$$

are based on two $p \times p$ matrices of random variables

$$S_e = Y'EY \quad \text{and} \quad S_h = Y'HY, \tag{9}$$

called respectively the sum of products matrix due to error and the sum of products matrix due to hypothesis. Here E and H are $n \times n$ symmetric idempotent matrices with non-stochastic elements, E of rank $(n - r)$ and H of rank t, E being a function of A, and H of both A and B. The likelihood-ratio test [6] is

$$\text{accept } \mathcal{H}_0 \text{ if } L \equiv \frac{|S_e|}{|S_e + S_h|} > c, \tag{10}$$
$$\text{otherwise reject } \mathcal{H}_0,$$

where c is a preassigned constant depending on the level of significance. The test based on the largest latent root [3] is

$$\text{(11)} \qquad \begin{array}{l} \text{accept } \mathcal{H}_0 \text{ if } \lambda_{\max}(S_h S_e^{-1}) < d, \\ \text{otherwise reject } \mathcal{H}_0 , \end{array}$$

where d is a constant depending on the level of significance. Simultaneous confidence-bounds on certain meaningful parametric functions have been derived by the largest (or the largest-smallest roots) procedure, [3] [4], whereas no such bounds are available as of now from the likelihood-ratio procedure.

2.3 *The step-down procedure*. We shall denote the ith columns of the matrices Y and Θ in section 2.2 by y_i and θ_i respectively and write $Y_i = [y_1\ y_2\ \cdots\ y_i]$ and $\Theta_i = [\theta_1\ \theta_2\ \cdots\ \theta_i]$. Further, we shall denote the top left-hand $i \times i$ submatrix of $\Sigma \equiv ((\sigma_{ij}))$ by Σ_i.

Then, under the condition that Y_i is fixed, the n elements of the vector y_{i+1} are distributed independently and normally each with the same variance σ_{i+1}^2 and expectations given by

$$\text{(12)} \qquad \mathcal{E}y_{i+1} = A\eta_{i+1} + Y_i\beta_i ,$$

where β_i is a vector of the form $i \times 1$ given by

$$\text{(13)} \qquad \beta_i = \Sigma_i^{-1} \begin{bmatrix} \sigma_{1,i+1} \\ \sigma_{2,i+1} \\ \cdots\cdots \\ \sigma_{i,i+1} \end{bmatrix}, \qquad \beta_0 = 0,$$

amd η_{i+1} is a vector of the form $m \times 1$ given by

$$\text{(14)} \qquad \eta_{i+1} = \theta_{i+1} - \Theta_i\beta_i$$

and

$$\text{(15)} \qquad \sigma_{i+1}^2 = \frac{|\Sigma_{i+1}|}{|\Sigma_i|},$$

with the understanding that $|\Sigma_0| = 1$ so that $\sigma_1^2 = \sigma_{11}$, $i = 0, 1, 2, \cdots, (p-1)$. The elements of the vectors β_i, η_{i+1} may then be regarded as unknown parameters. We shall call β_i the ith order step-down regression coefficient and σ_{i+1}^2 the ith order step-down residual variance.

Let us now consider linear functions

$$\text{(16)} \qquad \phi_i = B\eta_i \qquad (i = 1, 2, \cdots, p).$$

If Y_i is fixed, (12) is of the same form as (1). Let us now, with an easily understood notation similar to that used in Section 2.1, construct the statistics

$$\text{(17)} \qquad F_i \equiv \frac{(\hat{\phi}_i - \phi_i)' C_i^{-1} (\hat{\phi}_i - \phi_i)'/t}{s_i^2/(n - r - i + 1)} \qquad (i = 1, 2, \cdots, p).$$

Obviously, when Y_{i-1} is fixed, the statistic F_i is distributed as a variance ratio with t and $(n - r - i + 1)$ degrees of freedom $(i = 2, 3, \cdots, p)$. Finally, we note that in its functional form F_i involves only Y_i $(i = 1, 2, \cdots, p$ and that the conditional distribution of F_i, given Y_{i-1} does not involve Y_{i-1} $(i = 2, 3, \cdots, p)$ and hence $F_{i-1}, \cdots, F_1$. Also, F_1 is marginally distributed as a variance-ratio with t and $(n - r)$ degrees of freedom. Therefore the statistics $F_1, F_2, \cdots, F_p$ are independent. This can be verified in a straight-forward manner by using the transformation to rectangular coordinates as in [5] or any other set of step-down variates, or even otherwise.

For a preassigned constant α_i, $0 < \alpha_i < 1$, let f_i denote the upper $100\alpha_i$ per cent point of the variance-ratio distribution with t and $(n - r - i + 1)$ degrees of freedom. Then the probability P that simultaneously

$$F_i \leqq f_i, \qquad i = 1, 2, \cdots, p, \tag{18}$$

is given by

$$P = \prod_{i=1}^{p} (1 - \alpha_i). \tag{19}$$

Therefore, for any subset T of the natural numbers $1, 2, \cdots, t$ writing as in (6), $T[\boldsymbol{\phi}_i]$ and $T[\hat{\boldsymbol{\phi}}_i]$ for the T-norms of $\boldsymbol{\phi}_i$ and $\hat{\boldsymbol{\phi}}_i$ respectively, and setting

$$f_i'^2 = tf_i/(n - r - i + 1) \tag{20}$$

and writing $C_{i(T)}$ for the T-submatrix of C_i, we have the simultaneous confidence bounds

$$T[\hat{\boldsymbol{\phi}}_i] - f_i' s_i \lambda_{\max}^{1/2}(C_{i(T)}) \leqq T[\boldsymbol{\phi}_i] \leqq T[\hat{\boldsymbol{\phi}}_i] + f_i' s_i \lambda_{\max}^{1/2}(C_{i(T)}) \tag{21}$$

for all subsets T of $(1, 2, \cdots, t)$ and $i = 1, 2, \cdots, p$ with probability greater than P.

To derive a test of the hypothesis $\mathcal{H}_0$ that $\Phi = 0$, we note that $\mathcal{H}_0$ is true if and only if the hypothesis $\mathcal{H}_i$ that $\boldsymbol{\phi}_i = 0$ holds for all $i = 1, 2, \cdots, p$. Using the result (17), we set up the following procedure for testing $\mathcal{H}_0$:

$$\text{(22)}\quad \begin{array}{l} \text{accept } \mathcal{H}_0 \text{ if } u_i \equiv \dfrac{\hat{\boldsymbol{\phi}}_i' C_i^{-1} \hat{\boldsymbol{\phi}}_i / t}{s_i^2/(n - r - i + 1)} \leqq f_i \qquad \text{for all } i = 1, 2, \cdots, p; \\ \text{otherwise reject } \mathcal{H}_0 . \end{array}$$

Obviously, the level of significance for this test is $1 - P$ where P is given by (19). The arbitrariness in determining the f_i's when the level of significance is preassigned may be removed by stipulating that $\alpha_1 = \alpha_2 = \cdots = \alpha_p$. From the fact that the variance-ratio test (7) is uniformly unbiassed, it can be seen after a little consideration, that the test procedure (22) is also uniformly unbiassed.

To carry out the test one should first compute u_1. If $u_1 > f_1$, $\mathcal{H}_0$ is rejected and no further computations are needed. If $u_1 \leqq f_1$, the next step is to compute u_2. If $u_2 > f_2$, $\mathcal{H}_0$ is rejected and no further computations are needed. If $u_2 \leqq f_2$,

one proceeds to compute u_3 and so on. This way one need compute u_i if and only if $u_j \leqq f_j$ for $j = 1, 2, \cdots, i-1$. Much computational labor is saved thereby.

It is well known that the likelihood-ratio statistic L given by (10) can be expressed as

$$L = \prod_{i=1}^{p} \frac{(n-r-i+1)}{t+(n-r-i+1)u_i} \tag{23}$$

and this has been utilized [1] to obtain the moments of L when $\mathcal{H}_0$ is true. However, the step-down procedure based on the individual u_i's rather than on a single function L, is advantageous from the point of view of (i) setting up simultaneous confidence bounds and (ii) saving computational labor, specially in the situation indicated in the introduction.

3. Step-down procedure for variance-covariance matrices. Let $S_{p \times p} \equiv ((s_{ij}))$ be a symmetric matrix of random variables, distributed in Wishart's form with n degrees of freedom, $n > p$, so that S/n provides an unbiassed estimate for the variance-covariance matrix Σ of a p-variate normal population. In the same way as in Section 2.3, we shall write S_i for the $i \times i$ top left-hand submatrix of S and let

$$b_i = S_i^{-1} \begin{bmatrix} s_{1,i+1} \\ s_{2,i+1} \\ \cdots \\ s_{i,i+1} \end{bmatrix}, \qquad b_0 = 0, \tag{24}$$

$$s_{i+1}^2 = \frac{|S_{i+1}|}{|S_i|}, \qquad s_1^2 = s_{11}, \tag{25}$$

for $i = 1, 2, \cdots, p-1$. Let β_{i-1} and σ_i^2 be defined by (13) and (15) for $i = 1, 2, \cdots, p$. Then it is well known that when S_i is fixed, the distribution of b_i is independent of the distribution of s_{i+1}^2; the distribution of b_i is i-variate normal with expectation β_i and variance-covariance matrix $\sigma_{i+1}^2 S_i^{-1}$, and s_{i+1}^2/σ_{i+1}^2 has the chi-square distribution with $(n-i)$ degrees of freedom, $i = 1, 2, \cdots, (p-1)$. Finally s_1^2/σ_1^2 has the chi-square distribution with n degrees of freedom.

When more than one variance-covariance matrix is involved, we shall distinguish them by a superscript under parentheses. Thus with a number of population variance-covariance matrices $\Sigma^{(j)}$ and the corresponding Wishart matrices $S^{(j)}$, the quantities $\beta_i^{(j)}$, $\sigma_i^{(j)}$, $b_i^{(j)}$, $s_i^{(j)}$, etc., will be defined in the same way as in (13), (15), (24), and (25) for $j = 1, 2, \cdots$, etc.

3.1 *One variance-covariance matrix.* On the basis of a matrix S distributed in Wishart's form with n degrees of freedom, with S/n providing an unbiassed estimate for Σ, it is possible to set up simultaneous confidence-bounds on parameters which are functions of the elements of Σ by the step-down procedure as follows.

When S_i is fixed, the statistics $u = (b_i - \beta_i)'S_i(b_i - \beta_i)/\sigma_{i+1}^2$ and $v =$

s_{i+1}^2/σ_{i+1}^2 are distributed independently as chi-squares, u with i degrees of freedom and v with $n - i$ degrees of freedom. Therefore, given pre-assigned positive constants a_i, c_{i+1}, and d_{i+1}, where $c_{i+1} < d_{i+1}$, the probability P_{i+1} that

$$(b_i - \beta_i)'S_i(b_i - \beta_i)/s_{i+1}^2 \leqq a_i^2, \qquad c_{i+1} \leqq s_{i+1}^2/\sigma_{i+1}^2 \leqq d_{i+1} \tag{26}$$

holds for fixed S_i, is a constant depending only on n, i, a_i, c_{i+1}, and d_{i+1}. As a matter of fact,

$$P_{i+1} = \int_{c_{i+1}}^{d_{i+1}} G_i(a_i^2 x) g_{n-i}(x)\, dx \qquad (i = 1, 2, \cdots, p - 1), \tag{27}$$

where

$$G_\nu(x) = \int_0^x g_\nu(\xi)\, d\xi \tag{28}$$

and

$$g_\nu(x) = \frac{e^{-x} x^{\frac{1}{2}\nu - 1}}{2^{\frac{1}{2}\nu}\Gamma(\frac{1}{2}\nu)}. \tag{29}$$

Also, given preassigned positive constants b_1, $c_1(b_1, < c_1)$, the marginal probability P_1 that

$$c_1 \leqq s_1^2/\sigma_1^2 \leqq d_1 \tag{30}$$

is given by

$$P_1 = \int_{c_1}^{d_1} g_n(x)\, dx. \tag{31}$$

By an argument similar to that which follows (17) in section 2.3, we obtain the probability P that simultaneously

$$c_i \leqq s_i^2/\sigma_i^2 \leqq d_i \qquad (i = 1, 2, \cdots, p), \qquad (b_i - \beta_i)' S_i(b_i - \beta_i)/s_{i+1}^2 \leqq a_i^2 \quad (i = 1, 2, \cdots, p - 1) \tag{32}$$

as

$$P = \prod_{i=1}^{p} P_i.$$

Now, as in Section 2.3, for a given subset T_i of the integers $1, 2, \cdots, i$, writing $T_i[\beta_i]$ and $T_i[b_i]$ for the T_i-norms of β_i and b_i respectively, and writing $U_{i(T_i)}$ for the T_i-submatrix of S_i^{-1},

$$s_i^2/d_i \leqq \sigma_i^2 \leqq s_i^2/c_i \qquad \text{for } i = 1, 2, \cdots, p, \qquad T_i[b_i] - a_i s_{i+1}\lambda_{\max}^{1/2}(U_{i(T_i)}) \leqq T_i[\beta_i] \leqq T_i[b_i] + a_i s_{i+1}\lambda_{\max}^{1/2}(U_{i(T_i)}) \tag{33}$$

for all subsets T_i of $(1, 2, \cdots, i)$ and $i = 1, 2, \cdots, p-1$. The statement (33) thus provides simultaneous confidence-bounds on p parameters of the type σ_i^2 and $(2^p - p)$ parameters of the form $T_i[\beta_i]$ with probability not less than P.

It is to be noted that to set up simultaneous confidence bounds of the type (32), one has to evaluate the integral (27) which is not usually available in tabulated form. Another meaningful procedure, which, incidentally, avoids this difficulty, is to set up separate sets of simultaneous confidence bounds: one on $\sigma_1^2, \cdots, \sigma_p^2$, using the chi-square distribution for s_i^2/σ_i^2, with a preassigned probability and another set on the step-down regressions β_i, using the variance-ratio distribution for $(b_i - \beta_i)'S_i(b_i - \beta_i)/s_{i+1}^2$, and with a probability not less than a preassigned level.

We suggest a slightly different procedure for testing the hypothesis $\mathcal{H}_0$ that Σ has a specified value Σ_0. This hypothesis may be reformulated in terms of the step-down regression-coefficients and residual variances as follows: the hypothesis $\mathcal{H}_0$ is true if and only if each of the hypotheses

$$\mathcal{H}_{i1} : \sigma_i^2 = \sigma_{i0}^2, \qquad i = 1, 2, \cdots, p,$$

$$\mathcal{H}_{i2} : \beta_i = \beta_{i0}, \qquad i = 1, 2, \cdots, p-1,$$

is true, where σ_{i0}^2, β_{i0} are derived from Σ_0 the same way as σ_i^2, β_i are derived from Σ. The test procedure suggested is:

accept $\mathcal{H}_0$ if

$$(34) \qquad \begin{aligned} c_i \leqq s_i^2/\sigma_{i0}^2 \leqq d_i \qquad & (i = 1, 2, \cdots, p), \\ (b_i - \beta_{i0})'S_i(b_i - \beta_{i0})/\sigma_{i+1,0}^2 \leqq e_i^2 \quad & (i = 1, 2, \cdots, p-1); \end{aligned}$$

otherwise reject $\mathcal{H}_0$.

The level of significance α for this procedure is given by

$$(35) \qquad \alpha = 1 - \left\{\prod_{i=1}^{p} P_i'\right\}\left\{\prod_{i=1}^{p-1} P_i''\right\},$$

where

$$P_i' = \int_{c_i}^{d_i} g_{n-i+1}(x)\, dx,$$

$$P_i'' = G_i(e_i^2).$$

For a given α, the c_i, d_i, e_i's are not uniquely determined. The arbitrariness may be removed, for instance, by the further stipulation that

$$P_1' = P_2' = \cdots = P_p' = P_1'' = P_2'' = \cdots = P_{p-1}'' = \beta \text{ (say)}$$

and that (c_i, d_i) are the locally unbiassed partitioning of the 100 $(1 - \beta)$ per cent critical region based on the chi-square distribution with $n - i + 1$ degrees of freedom. With this choice of the constants c_i, d_i, e_i, the test procedure is locally unbiassed.

3.2 *Two variance-covariance matrices.* With two population variance-covariance

matrices $\Sigma^{(1)}$, $\Sigma^{(2)}$ and two matrices of random variables $S^{(1)}$, $S^{(2)}$ distributed independently in Wishart's form with n_1 and n_2 degrees of freedom respectively, so that $S^{(j)}/n_j$ provides an unbiassed estimate for $\Sigma^{(j)}$, we can use the step-down procedure for testing the hypothesis $\mathcal{H}_0$ that the two variance-covariance matrices are identical or, in symbols,

$$\mathcal{H}_0 : \Sigma^{(1)} = \Sigma^{(2)},$$

and also set up simultaneous confidence bounds for parameters measuring deviations from $\mathcal{H}_0$.

Let us introduce the two sets of step-down regression-coefficients and residual variances: $\beta_i^{(j)}$, $\sigma_i^{(j)}$, $b_i^{(j)}$, and $s_i^{(j)}$. The hypothesis $\mathcal{H}_0$ may be reformulated in terms of the step-down parameters as follows: $\mathcal{H}_0$ is true if and only if the hypotheses

$$\text{(36)} \qquad \begin{aligned} &\mathcal{H}_{i1} : \sigma_i^{(1)} = \sigma_i^{(2)}, && i = 1, 2, \cdots, p, \\ &\mathcal{H}_{i2} : \beta_i^{(1)} = \beta_i^{(2)}, && i = 1, 2, \cdots, p-1, \end{aligned}$$

are simultaneously true. We may take $\rho_i = \sigma_i^{(1)}/\sigma_i^{(2)}$ and $T_i[\delta_i]$ as measures of deviation from $\mathcal{H}_0$ where $\delta_i = \beta_i^{(1)} - \beta_i^{(2)}$, T_i is a subset of $(1, 2, \cdots, i)$ and $T_i[\delta_i]$ denotes the T_i-norm of δ_i. In this case, it has not been possible to set-up confidence bounds on all these parameters simultaneously. However, one may proceed as follows. Given pre-assigned positive constants c_i, d_i; $c_i < d_i$, and writing

$$\text{(37)} \qquad r_i = \left(\frac{n_1 - i + 1}{n_2 - i + 1}\right)^{-1/2} s_i^{(1)}/s_i^{(2)},$$

we find the probability that

$$\text{(38)} \qquad r_i^2/d_i \leqq \rho_i^2 \leqq r_i^2/c_i, \qquad i = 1, 2, \cdots, p,$$

should hold simultaneously is given by

$$\text{(39)} \qquad P = \prod_{i=1}^{p} P_i,$$

where

$$\text{(40)} \qquad P_i = \int_{c_i}^{d_i} dF_{n_2-i+1}^{n_1-i+1}(x),$$

in which $F_n^m(x)$ stands for the distribution-function of the variance-ratio statistic with m degrees of freedom for the numerator and n degrees of freedom for the denominator. Therefore, (38) provides simultaneous confidence-bounds on ρ_i^2 $(i = 1, 2, \cdots, p)$ with probability P.

Let us now write $\hat{\delta}_i = b_i^{(1)} - b_i^{(2)}$ and note that if $S_i^{(1)}$ and $S_i^{(2)}$ are fixed, $\hat{\delta}_i$ is distributed in an i-variate normal form with expected value δ_i and variance-covariance matrix

$$\{\sigma_{i+1}^{(1)}\}^2\{S_i^{(1)}\}^{-1} + \{\sigma_{i+1}^{(2)}\}^2\{S_i^{(2)}\}^{-1}$$

distributed independently of $s_{i+1}^{(1)}$ and $s_{i+1}^{(2)}$. If $\mathcal{H}_{i+1,1}$ is true, we have $\sigma_{i+1}^{(1)} = \sigma_{i+1}^{(2)} = \sigma_{i+1}$, say. In that case, if $S_i^{(1)}$ and $S_i^{(2)}$ are fixed, $\hat{\delta}_i$ is distributed in an i-variate normal form with expected value δ_i and dispersion matrix $C_i . \sigma_{i+1}^2$ where

$$C_i = \{S_i^{(1)}\}^{-1} + \{S_i^{(2)}\}^{-1}. \tag{41}$$

Also, $\hat{\delta}_i$ is distributed independently of u_1 and u_2 where

$$u_j = (s_{i+1}^{(j)})^2/\sigma_{i+1}^2 \qquad (j = 1, 2) \tag{42}$$

and u_j is distributed as a chi-square with $(n_j - i)$ degrees of freedom. Consequently, writing

$$s_{i+1}^2 = (s_{i+1}^{(1)})^2 + (s_{i+1}^{(2)})^2 \tag{43}$$

we find that if $\mathcal{H}_{i+1,1}$ is true and $S_i^{(j)}$ are fixed $(j = 1, 2)$ the statistics

$$(\hat{\delta}_i - \delta_i)' C_i^{-1} (\hat{\delta}_i - \delta_i)/s_{i+1}^2 \tag{44}$$

and

$$\frac{n_2 - i}{n_1 - i} \left(\frac{s_{i+1}^{(1)}}{s_{i+1}^{(2)}} \right)^2 \tag{45}$$

are distributed independently as variance-ratios, (44) with i and $(n_1 + n_2 - 2i)$ degrees of freedom, and (45) with $(n_1 - i)$ and $(n_2 - i)$ degrees of freedom.

Therefore, given pre-assigned positive quantities e_i^2 the probability P' that

$$(\hat{\delta}_i - \delta_i)' C_i^{-1} (\hat{\delta}_i - \delta_i)/s_{i+1}^2 \leq e_i^2, \quad i = 1, 2, \cdots, p - 1, \tag{46}$$

should hold simultaneously is equal to

$$P' = \prod_{i=1}^{p-1} P_i', \tag{47}$$

where

$$P_i' = F_{n_1+n_2-2i}^{i}(e_i^2) \tag{48}$$

provided $\mathcal{H}_{i1}$ is true for $i = 2, 3, \cdots, p$. From (45), we get the following simultaneous confidence-bounds (49) on the T_i-norms of δ_i where T_i is a subset of $(1, 2, \cdots, i)$ (under the highly restrictive condition that $\mathcal{H}_{i1}$ is true) for $i = 2, 3, \cdots, p$:

$$T_i[\hat{\delta}_i] - e_i s_{i+1} \lambda_{\max}^{1/2}(C_{i(T_i)}) \leq T_i[\delta_i] \leq T_i[\hat{\delta}_i] + e_i s_{i+1} \lambda_{\max}^{1/2}(C_{i(T_i)}) \tag{49}$$

with probability not less than P', where $C_{i(T_i)}$ is the T_i-submatrix of C_i.

To test the hypothesis $\mathcal{H}_0$, the step-down procedure suggested is:

accept $\mathcal{H}_0$ if

$$\begin{aligned} &(\hat{\delta}_i - \delta_i)' C_i^{-1} (\hat{\delta}_i - \delta_i)/s_{i+1}^2 \leq e_i^2, \quad i = 1, 2 \cdots, p - 1, \\ &c_i \leq \frac{n_2 - i + 1}{n_1 - i + 1} \frac{s_i^{(1)}}{s_i^{(2)}} \leq d_i, \qquad i = 1, 2, \cdots, p, \end{aligned} \tag{50}$$

and, otherwise, reject $\mathcal{H}_0$,

where e_i^2, c_i, $d_i(c_i < d_i)$ are pre-assigned positive constants. The level of significance α is given by

$$\alpha = 1 - \left\{\prod_{i=1}^{p} P_i\right\}\left\{\prod_{i=1}^{p-1} P_i'\right\}, \tag{51}$$

where P_i is given by (40) and P_i' by (48). For a pre-assigned value of α, the constants c_i, d_i, e_i^2 are uniquely determined if we stipulate that

$$P_1 = P_2 = \cdots = P_p = P_1' = P_2' = \cdots = P_{p-1}' = \beta, \quad \text{say},$$

and that (c_i, d_i) gives an unbiassed partitioning of the $100(1 - \beta)$ per cent critical region of the variance-ratio distribution with i and $n_1 + n_2 - 2i$ degrees of freedom. With this choice the step-down test is locally unbiassed.

4. Acknowledgment. The author wishes to thank Professor S. N. Roy for kindly going through the manuscript and suggesting improvements.

REFERENCES

[1] C. R. Rao, "On transformations useful in the distribution problems of least squares," *Sankhyā*, Vol. 12 (1952-53), pp. 339-346.

[2] S. N. Roy, "On a heuristic method of test construction and its use in multivariate analysis," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 220-238.

[3] S. N. Roy, "A report on some aspects of multivariate analysis," North Carolina Institute of Statistics Mimeograph Series No. 121, 1954.

[4] S. N. Roy and R. C. Bose, "Simultaneous confidence interval estimation," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 513-536.

[5] S. N. Roy and R. E. Bargmann, "Tests of multiple independence and the associated confidence-bounds," North Carolina Institute of Statistics Mimeograph Series No. 175, 1957.

[6] S. S. Wilks, "Certain generalizations in the analysis of variance," *Biometrika*, Vol. 24 (1932), pp. 471-494.

THE LIMITING DISTRIBUTION OF THE SERIAL CORRELATION COEFFICIENT IN THE EXPLOSIVE CASE

By John S. White

Aero Division, Minneapolis Honeywell Regulator Company, Minneapolis Minnesota

1. Introduction and summary. Several authors have studied the discrete stochastic process (x_t) in which the x's are related by the stochastic difference equation

$$x_t = \alpha x_{t-1} + u_t, \qquad t = 1, 2, \cdots, T, \tag{1.1}$$

where the u's are unobservable disturbances, independent and identically distributed with mean zero and variance σ^2, and α is an unknown parameter.

The statistical problem is to find some appropriate function of the x's as an estimator for α and examine its properties.

We may rewrite (1.1) as

$$x_t = u_t + \alpha u_{t-1} + \cdots + \alpha^{t-1} u_1 + \alpha^t x_0. \tag{1.2}$$

From (1.2) we see that the distribution of the successive x's is not uniquely determined by that of the u's alone. The distribution of x_0 must also be specified. Three distributions which have been proposed for x_0 are the following:

(A) x_0 = a constant (with probability one),

(B) x_0 is normally distributed with mean zero and variance $\sigma^2/(1 - \alpha^2)$,

(C) $x_0 = x_T$.

Distribution (B) is perhaps the most appealing from a physical point of view, since if x_0 has this distribution and if the u's are normally distributed, then the process is stationary (e.g., see Koopmans [4]). However, there are several analytic difficulties which arise in the statistical treatment of this process. Distribution (C), the so-called circular distribution, has been proposed as an approximation to (B) and is much easier to analyze (e.g., see Dixon [2]). Distribution (A) has been studied extensively by Mann and Wald [5]. An interesting feature of distribution (A) is that α may assume any finite value, while for distributions (B) and (C) α must be between -1 and 1. From (1.2) we see that a process satisfying (1.1) and (A) has

$$\text{var}(x_t) = \sigma^2(1 + \alpha^2 + \cdots + \alpha^{2(t-1)}). \tag{1.3}$$

If $|\alpha| \geqq 1$, $\lim_{t\to\infty} \text{var}(x_t) = \infty$ and the process is said to be "explosive."

Mann and Wald [5] considered only the case $|\alpha| < 1$. They showed that the least squares estimator for α is the serial correlation coefficient[1]

$$\hat{\alpha} = \frac{\sum x_t x_{t-1}}{\sum x_{t-1}^2} \tag{1.4}$$

Received December 30, 1955; revised May 27, 1958.

[1] In this paper, the summation sign $\sum$ will always refer to summation from $t = 1$ to $t = T$.

and that (for $|\alpha| < 1$) this estimator is asymptotically normally distributed with mean α and variance $(1 - \alpha^2)/T$. Rubin [6] showed that the estimator $\hat{\alpha}$ is consistent (i.e., plim $\hat{\alpha} = \alpha$) for all α.

In this paper the asymptotic distribution of $\hat{\alpha}$ will be studied under the assumption that the u's are normally distributed. For $|\alpha| > 1$, it is shown that the asymptotic distribution of α is the Cauchy distribution. For $|\alpha| = 1$, a moment generating function is found, the inversion of which will yield the asymptotic distribution.

2. The distribution of $\hat{\alpha} - \alpha$. From equation (1.1) and condition (A) the joint distribution of

$$x' = (x_1, x_2, \cdots, x_T)$$

is easily found to be

$$f(x') = \frac{\exp\left[(-1/2\sigma^2) \sum (x_t - \alpha x_{t-1})^2\right]}{(2\pi\sigma^2)^{T/2}}. \tag{2.1}$$

The maximum likelihood estimator for α is then the least-squares estimator $\hat{\alpha}$. Since we shall be considering only the distribution of

$$\hat{\alpha} = \frac{\sum x_t x_{t-1}}{\sum x_{t-1}^2},$$

we may, without loss of generality, take $\sigma^2 = 1$. For the time being we shall also set $x_0 = 0$.

We may now write (2.1) in matrix form as follows:

$$f(x') = \frac{\exp(-\frac{1}{2}x'Px)}{(2\pi)^{T/2}}, \tag{2.2}$$

where P is the $T \times T$ matrix

$$P = \begin{bmatrix} 1+\alpha^2 & -\alpha & 0 & 0 & & & \\ -\alpha & 1+\alpha^2 & -\alpha & 0 & & & \\ 0 & -\alpha & 1+\alpha^2 & -\alpha & & & \\ & & & & \cdots & & \\ & & & & -\alpha & 1+\alpha^2 & -\alpha \\ & & & & 0 & -\alpha & 1 \end{bmatrix}. \tag{2.3}$$

Since $\hat{\alpha}$ is a consistent estimator for α, we shall consider the distribution of $\hat{\alpha} - \alpha$ rather than that of α alone. We have

$$\begin{aligned} \hat{\alpha} - \alpha &= \frac{\sum x_t x_{t-1}}{\sum x_{t-1}^2} - \alpha \\ &= \frac{\sum x_t x_{t-1} - \alpha \sum x_{t-1}^2}{\sum x_{t-1}^2} \\ &= \frac{x'Ax}{x'Bx}, \end{aligned} \tag{2.4}$$

where A and B are the $T \times T$ matrices

$$A = -\tfrac{1}{2}\begin{bmatrix} 2\alpha & -1 & 0 & & & & \\ -1 & 2\alpha & -1 & & & & \\ 0 & -1 & 2\alpha & & & & \\ & & & \cdots & & & \\ & & & & -1 & 2\alpha & -1 \\ & & & & 0 & -1 & 0 \end{bmatrix},$$

$$(2.5)$$

$$B = \begin{bmatrix} 1 & 0 & 0 & & & & \\ 0 & 1 & 0 & & & & \\ 0 & 0 & 1 & & & & \\ & & & \cdots & & & \\ & & & & 0 & 1 & 0 \\ & & & & 0 & 0 & 0 \end{bmatrix}.$$

Let $m(u, v)$ be the joint moment generating function of $x'Ax$ and $x'Bx$. We have

$$\begin{aligned} m(u, v) &= E\,(\exp\{x'Axu + x'Bxv\}) \\ &= (2\pi)^{-T/2}\int \exp\,(x'Axu + x'Bxv - x'Px/2)\,dx \qquad (2.6) \\ &= (2\pi)^{-T/2}\int \exp\,(-x'Dx/2)\,dx, \end{aligned}$$

where D is the $T \times T$ matrix

$$D = P - 2Au - 2Bv = \begin{bmatrix} p & q & 0 & & & & \\ q & p & q & & & & \\ 0 & q & p & & & & \\ & & & \cdots & & & \\ & & & & q & p & q \\ & & & & 0 & q & 1 \end{bmatrix}, \qquad (2.7)$$

$$p = 1 + \alpha^2 - 2v + 2\alpha u, \qquad q = -(\alpha + u).$$

By a well-known integration formula (Cramer [1], Eq. (11.12.2.), p. 120) we have

$$m(u, v) = (2\pi)^{-T/2}\int \exp\left(-\frac{x'Dx}{2}\right)dx = (\det D)^{-\frac{1}{2}}. \qquad (2.8)$$

If we now write $\det D = D(T)$, we note that expanding (2.7) by the elements of the first column gives the difference equation

$$D(T) = pD(T-1) - q^2D(T-2). \qquad (2.9)$$

From the initial values $D(1) = 1$ and $D(2) = p - q^2$, we obtain

$$D(T) = \frac{1-s}{r-s}\,r^T + \frac{1-r}{s-r}\,s^T, \qquad (2.10)$$

where r and s are roots of the equation $x^2 - px + q^2 = 0$, that is

$$r, s = (p \pm \sqrt{p^2 - 4q^2})/2. \tag{2.11}$$

The inversion of $m(u, v) = D(T)^{-\frac{1}{2}}$ seems out of the question for finite T. The inversion of a certain limiting form of $m(u, v)$ will be discussed in Section 4.

3. The standardizing function $g(T)$. Since $\hat{\alpha}$ is consistent the limiting distribution of $\hat{\alpha} - \alpha$ is the unitary distribution. The first problem then is to find some function of T, say $g(T)$, such that the limiting distribution of $g(T)$ $(\hat{\alpha} - \alpha)$ is non-degenerate. We note that the results of Mann and Wald (Eq. (1.4) above) give $g(T) = (T/\{1 - \alpha^2\})^{\frac{1}{2}}$ for $|\alpha| < 1$, since $(T/\{1 - \alpha^2\})^{\frac{1}{2}}$ $(\hat{\alpha} - \alpha)$ has a limiting normal distribution. The function $g^2(T)$ corresponds roughly to the reciprocal of the asymptotic variance of $(\hat{\alpha} - \alpha)$, or in Fisher's terminology the "information" on α supplied by the sample.

The "information" on α may be obtained explicitly as follows. Let f be the density function (2.1) with $x_0 = 0$ and $\sigma^2 = 1$. The "information," say $I(\alpha)$, is then defined as

$$\begin{aligned} I(\alpha) &= E\left(-\frac{d^2 \log f}{d\alpha^2}\right) \\ &= E\left(\sum x_{i-1}^2\right) \\ &= \frac{1}{1-\alpha^2}\left(T - \frac{1-\alpha^{2T}}{1-\alpha^2}\right) \quad \text{if} \quad |\alpha| \neq 1 \\ &= \frac{T(T-1)}{2} \quad \text{if} \quad |\alpha| = 1. \end{aligned} \tag{3.1}$$

If the x's had been independent random variables, then $I(\alpha)$ $(\hat{\alpha} - \alpha)$ would be asymptotically N (0, 1) (Cramer [1], Eq.(33.3.4), p. 503). This, of course, is not the case. This approach does, however, give an heuristic method for finding a function $g(T)$ such that $g(T)$ $(\hat{\alpha} - \alpha)$ has a non-degenerate limiting distribution.

We might now take $g(T) = [I(\alpha)]^{\frac{1}{2}}$; however, it will simplify the computations to use slight modifications which are asymptotically equivalent to $[I(\alpha)]^{\frac{1}{2}}$. We choose

$$\begin{aligned} g(T) &= \sqrt{\frac{T}{1-\alpha^2}} \qquad \text{for } |\alpha| < 1, \\ &= \frac{T}{\sqrt{2}} \qquad \text{for } |\alpha| = 1, \\ &= \frac{|\alpha|^T}{\alpha^2 - 1} \qquad \text{for } |\alpha| > 1. \end{aligned} \tag{3.2}$$

In the next section it will be shown that $g(T)$ $(\hat{\alpha} - \alpha)$ has a non-degenerate distribution for all values of α.

4. The limiting distribution of $g(T)$ $(\hat{\alpha} - \alpha)$. We shall first consider the joint distribution of $x'Ax/g(T)$ and $x'Bx/g^2(T)$. Let $M(U, V)$ be the joint moment generating function of these two statistics. We then have

$$
\begin{aligned}
M(U, V) &= E[\exp x'AxU/g(T) + x'BxV/g^2(T)] \\
&= m[U/g(T), V/g^2(T)],
\end{aligned} \tag{4.1}
$$

where $m(u, v)$ is the joint moment generating function (2.6).

From (2.10) and (2.11) with $g = g(T)$, $u = U/g$ and $v = V/g^2$, we have

$$
\begin{aligned}
M(U, V) &= D(T)^{-1/2} \\
&= \frac{1 - s}{r - s} r^T + \frac{1 - r}{s - r} s^T,
\end{aligned} \tag{4.2}
$$

$$
\begin{aligned}
r, s = \tfrac{1}{2}[1 + \alpha^2 + 2\alpha U/g - 2V/g^2 \pm \{(1 - \alpha^2)^2 - 4\alpha(1 - \alpha^2)U/g \\
- 4(1 - \alpha^2)U^2/g^2 - 4(1 + \alpha^2)V/g^2 - 8\alpha UV/g^3 + 4V^2/g^4\}^{1/2}].
\end{aligned} \tag{4.3}
$$

For sufficiently large T and $|\alpha| \neq 1$, we may factor $(1 - \alpha^2)$ out of the radical in (4.3) and expand the remaining radical by the binomial theorem. We then have, up to terms of order $O(g^{-3})$

$$
\begin{aligned}
r, s = \frac{1}{2}\Bigg[& 1 + \alpha^2 + 2\alpha U/g - 2V/g^2 \\
& \pm \left\{ 1 - \alpha^2 - 2\alpha U/g - \frac{2(1 + \alpha^2)V}{(1 - \alpha^2)g^2} - \frac{2U^2}{(1 - \alpha^2)g^2} + O(g^{-3}) \right\} \Bigg].
\end{aligned} \tag{4.4}
$$

Taking r with the plus sign and s with the minus sign we have

$$
\begin{aligned}
r &= 1 - \frac{U^2 + 2V}{(1 - \alpha^2)g^2} + O(g^{-3}), \\
s &= \alpha^2 + 2\alpha U/g + \frac{U^2 + 2\alpha^2 V}{(1 - \alpha^2)g^2} + O(g^{-3}).
\end{aligned} \tag{4.5}
$$

Substituting the appropriate values of $g(T)$ from (3.2), we have

$$
\begin{aligned}
r &= 1 - \frac{U^2 + 2V}{T} + O(T^{-3/2}) \qquad \text{for } |\alpha| < 1, \\
s &= \alpha^2 + 2\alpha \sqrt{\frac{1 - \alpha^2}{T}}\, U + \frac{U^2 + 2\alpha^2 V}{T} + O(T^{-3/2}).
\end{aligned} \tag{4.6}
$$

$$
\begin{aligned}
r &= 1 + \frac{(U^2 + 2V)(\alpha^2 - 1)}{\alpha^{2T}} + O(|\alpha|^{-3T}) \qquad \text{for } |\alpha| > 1, \\
s &= \alpha^2 + \frac{2\alpha U(\alpha^2 - 1)}{|\alpha|^T} - \frac{(U^2 + 2\alpha^2 V)(\alpha^2 - 1)}{\alpha^{2T}} + O(|\alpha|^{-3T}).
\end{aligned} \tag{4.7}
$$

If $|\alpha| = 1$, the expansion in (4.4) is not valid; however, from (4.3), we have

$$(4.8)\qquad \begin{aligned} r &= 1 + \frac{\sqrt{2}\alpha U}{T} + \frac{2i\sqrt{V}}{T} + O(T^{-2}) \qquad \text{for } |\alpha| = 1, \\ s &= 1 + \frac{\sqrt{2}\alpha U}{T} - \frac{2i\sqrt{V}}{T} + O(T^{-2}). \end{aligned}$$

Substituting these results in (4.2), we have

$$(4.9)\qquad \begin{aligned} \lim M(U, V) &= \exp\,(V + U^2/2) \qquad \text{for } |\alpha| < 1, \\ &= (1 - U^2 - 2V)^{-1/2} \qquad \text{for } |\alpha| > 1, \\ &= \exp\,(\sqrt{2}\alpha U)\left(\cos 2\sqrt{V} - \frac{\sqrt{2}\alpha U}{2\sqrt{V}} \sin 2\sqrt{V}\right)^{-\frac{1}{2}} \\ &\qquad\qquad \text{for } |\alpha| = 1. \end{aligned}$$

The next problem is to obtain the limiting distribution of $g(T)(\hat{\alpha} - \alpha)$ from $\lim M(U, V)$. Since $g(T)(\hat{\alpha} - \alpha) = g(T)x'Ax/x'bX$, the problem is one of finding the distribution of the ratio of two random variables. One method of solution has been proposed by Gurland [3]. Let X and Y be two random variables, Prob $(Y > 0) = 1$. We wish to determine the distribution of $Z = X/Y$. Let $W = W_z = X - zY$. Then we have

$$(4.10)\qquad \begin{aligned} \text{Prob}\,(Z < z) &= \text{Prob}\,(X/Y < z) \\ &= \text{Prob}\,(X - zY < 0) \\ &= \text{Prob}\,(W_z < 0). \end{aligned}$$

If the distribution of W can be found, the distribution of Z will immediately follow. Frequently the distribution of W can be found from that of X and Y by means of moment generating functions. Let

$$(4.11)\qquad m(w) = E(\exp\{Ww\}), \qquad m^*(u, v) = E(\exp\{Xu + Yv\}),$$

then

$$m(w) = E(\exp\{X - zY\}w) = E(\exp\{Xw - Yzw\}) = m^*(w, -zw).$$

To apply this technique to the problem at hand, we set $W = x'Ax/g - zx'Bx/g^2$. From (4.1), (4.2) and (4.9) we have

$$(4.12)\qquad \begin{aligned} m(w) &= M(w, -zw), \\ \lim m(w) &= \exp\,(-zw + w^2/2) \qquad \text{for } |\alpha| < 1, \\ &= (1 + 2zw - w^2)^{-\frac{1}{2}} \qquad \text{for } |\alpha| > 1, \\ &= \left\{\exp\,(\sqrt{2}\alpha w)\left(\cos 2\sqrt{-zw} - \frac{\sqrt{2}\alpha w}{2\sqrt{-zw}} \sin 2\sqrt{-zw}\right)\right\}^{-1/2} \\ &\qquad\qquad \text{for } |\alpha| = 1. \end{aligned}$$

The inversion of $\lim m(w)$ is trivial for $|\alpha| < 1$. The moment generating function $\exp(-zw+w^2/2)$ is immediately recognized as that of a random variable which is normally distributed with mean $-z$ and variance 1. Hence we have

$$
\begin{aligned}
\lim \text{Prob}\,(W < 0) &= (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{0} \exp\,(-\{t + z\}^2/2)\,dt \\
&= (2\pi)^{-1/2} \int_{-\infty}^{z} \exp\,(-t^2/2)\,dt \qquad (4.13)\\
&= \lim \text{Prob}\,\{g(T)(\hat{\alpha} - \alpha) < z\},
\end{aligned}
$$

i.e., $g(T)\,(\hat{\alpha} - \alpha)$ is asymptotically normal with mean 0 and variance 1.

For $|\alpha| > 1$, the inverse of $\lim m(w)$ might be obtained directly in terms of Bessel functions; however, it is more appealing from a statistical point of view to proceed as follows. Let X and Y be independent chi-squared variables with one degree of freedom. Then $E(\exp\{Xw\}) = E(\exp\{Yw\}) = (1 - 2w)^{-1/2}$ is their common moment generating function. Now set $R = aX - bY$, the moment generating function of R will be

$$
\begin{aligned}
m_R(w) = E(\exp\{Rw\}) &= E(\exp\{aX - bY\}w) \\
&= (\{1 - 2aw\}\{1 + 2bw\})^{-1/2}. \qquad (4.14)
\end{aligned}
$$

In particular if we set

$$
2a = \sqrt{1 + z^2} - z, \qquad 2b = \sqrt{1 + z^2} + z, \qquad (4.15)
$$

we have

$$
m_R(w) = (1 + 2zw - w^2)^{1/2} = \lim m(w). \qquad (4.16)
$$

Hence, the limiting distribution of W, for $|\alpha| > 1$, is the same as the distribution of $R = aX - bY$. We then have

$$
\begin{aligned}
\lim \text{Prob}\,(W < 0) &= \text{Prob}\,(aX - bY < 0) \\
&= \text{Prob}\,(X < bY/a) \\
&= \frac{1}{2\pi}\int_0^\infty \int_0^{by/a} \frac{\exp\,(-x/2 - y/2)}{\sqrt{xy}}\,dx\,dy \qquad (4.17)\\
&= \lim \text{Prob}\,\{g(T)(\hat{\alpha} - \alpha) < z\} = \text{say } F(z).
\end{aligned}
$$

The density function corresponding to $F(z)$ is

$$
\begin{aligned}
f(z) = \frac{dF(z)}{dz} &= \frac{1}{2\pi}\int_0^\infty \sqrt{a/b}\,\exp\,(-by/2a - y/2)\left\{\frac{d(b/a)}{dz}\right\} dy \\
&= \frac{1}{2\pi}\sqrt{a/b}\,\frac{2}{1 + (b/a)}\,\frac{d(b/a)}{dz} \qquad (4.18)\\
&= \frac{1}{\pi}\,\frac{1}{1 + z^2} \qquad \text{(by (4.15))}.
\end{aligned}
$$

Hence the limiting distribution of $g(T)(\hat{\alpha} - \alpha)$, for $|\alpha| > 1$, is the Cauchy distribution.

We have been unable to invert $\lim m(w)$ when $|\alpha| = 1$. In the next section certain results concerning this limit and more general problems of this type will be discussed.

If we now let $x_0 = c$, a non-zero constant, the analysis proceeds much as before. Let A, B, P, and D be the $T \times T$ matrices defined in (2.3), (2.5) and (2.7). We then have, analogous to (2.1) and (2.4),

$$(4.19)\qquad \begin{aligned} f(x') &= (2\pi)^{-T/2} \exp\,(cx_1\alpha - \alpha^2c^2/2 - x'Px/2), \\ \hat{\alpha} - \alpha &= \frac{x'Ax + cx_1 - \alpha c^2}{x'Bx + c^2}. \end{aligned}$$

The joint moment generating function of $x'Ax + cx_1 - \alpha^2c^2$ and $x'Bx + c^2$ is

$$(4.20)\qquad \begin{aligned} m(u, v) &= E\,(\exp\,\{(x'Ax + cx_1 - \alpha c^2)u + (x'Bx + c^2)v\}) \\ &= \left\{\exp\left(c^2v - c^2\alpha u - \frac{\alpha^2c^2}{2}\right)\right\}(2\pi)^{-T/2} \\ &\qquad \cdot \int \exp\left([u + \alpha]cx_1 - \frac{x'Dx}{2}\right) dx \\ &= \exp\left(c^2v - c^2\,\alpha u - \frac{\alpha^2c^2}{2}\right)\exp\left\{(u + \alpha)^2\frac{c^2}{2}\,\frac{D(T-1)}{D(T)}\right\} D(T)^{-1/2}, \end{aligned}$$

$$(4.21)\qquad \begin{aligned} \lim m(U/g, V/g^2) &= \lim M(U, V) \\ &= \lim\left\{D(T)^{-\frac{1}{2}} \exp\left(\frac{-\alpha^2c^2}{2}\left[1 - \frac{D'(T-1)}{D(T)}\right]\right)\right\}, \end{aligned}$$

where $D(T)$ is as defined in (4.2) while $D'(T-1)$ is defined in a similar fashion but with $g = g(T)$.

For $|\alpha| \leqq 1$, it follows from (4.6) and (4.8) that, since $g(T)$ and $g(T-1)$ are of the same order,

$$\lim D(T) = \lim D'(T-1)$$

and hence

$$(4.22)\qquad \lim m(U/g, V/g^2) = \lim M(U, V) = \lim D(T)^{-1/2}.$$

We see that this limit is the same as that for $x_0 = 0$ as given in (4.9) and hence the limiting distribution of $g(T)(\hat{\alpha} - \alpha)$ does not depend on the initial value x_0 for $|\alpha| \leqq 1$.

For $|\alpha| > 1$ we have, from (4.7),

$$(4.23)\qquad \begin{aligned} \lim D(T) &= 1 - (U + 2V), \\ \lim D'(T-1) &= \frac{(U + 2V)}{\alpha^2}; \end{aligned}$$

and in place of (4.22) we have

$$\begin{aligned}\lim M(U, V) &= \lim D(T)^{-1/2} \exp\left(-\frac{\alpha^2 c^2}{2}\left[1 - \frac{D'(T-1)}{D(T)}\right]\right) \\ &= (1 - U^2 - 2V)^{-1/2} \exp\left\{\frac{(\alpha^2 - 1)c^2}{2}\left(\frac{U^2 + 2V}{1 - U^2 - 2V}\right)\right\}.\end{aligned} \tag{4.24}$$

This moment generating function may be inverted by the methods of Section 4 to give

$$f(x) = \frac{e^{-q}}{\sqrt{\pi}(1 + x^2)} \sum_{k=0}^{\infty} \left(\frac{q}{1 + x^2}\right)^k \frac{1}{\Gamma(k + \frac{1}{2})}, \qquad q = \frac{c^2(\alpha^2 - 1)}{2}, \tag{4.25}$$

as the limiting distribution of $g(T)(\hat{\alpha} - \alpha)$. We note that for $c = 0$, $f(x)$ is the Cauchy distribution as obtained in (4.18).

6. Final remarks. The results of Mann and Wald [5] show that the limiting distribution of $g(T)(\hat{\alpha} - \alpha)$, for $|\alpha| < 1$, is also $N(0, 1)$ if, rather than assuming that the "errors" u_t are normally distributed, we merely assume that all of the moments of the u's are finite. This is another example of an invariance principle which seems to hold quite generally for the limiting distributions of function of random variables. Roughly speaking, there seems to be an unproved (and unstated) theorem that the limiting distribution of a function of a sequence of independent random variables, with suitable restrictions on these random variables, depends only on the form of the function and is the same as the distribution of a related functional on a stochastic process.

A general result of this form is Donsker's Theorem [7] which gives the limiting distribution of any function of sums of independent identically distributed random variables with finite variances as the distribution of a corresponding functional on the Wiener process. It is conjectured that this type of reasoning will show that the results of Mann and Wald will still hold if the u's are merely assumed to have finite variances.

For $\alpha = 1$, application of Donsker's Theorem shows that the limiting distribution of $g(T)(\hat{\alpha} - \alpha)$ is the same as the distribution of the functional

$$G[x(\cdot)] = \frac{\int_0^1 x(t)\, dx(t)}{\int_0^1 x^2(t)\, dt} = \frac{\frac{1}{2}x^2(1) - \frac{1}{2}}{\int_0^1 x^2(t)\, dt}$$

on the Wiener process, independent of the distribution of the u's. This distribution will be considered in a future paper.

REFERENCES

[1] H. Cramer, *Mathematical Methods of Statistics*, Princeton University Press, 1946.

[2] W. J. Dixon, "Further contributions to the problem of serial correlation," *Ann. Math. Stat.*, Vol. 15 (1944), pp. 119–144.

[3] J. GURLAND, "Inversion formulae for the distribution of ratios," *Ann. Math. Stat.*, Vol. 19 (1948), pp. 228-237.

[4] T. KOOPMANS, "Serial correlation and quadratic forms in normal variables," *Ann. Math. Stat.*, Vol. 13 (1942), pp. 1-42.

[5] H. B. MANN AND A. WALD, "On the statistical treatment of linear stochastic difference equations," *Econometrica*, Vol. 11 (1943), pp. 173-200.

[6] H. RUBIN, "Consistency of maximum likelihood estimates in the explosive case," *Statistical Inference in Dynamic Economic Models*, T. Koopmans (ed.), John Wiley & Sons, 1950.

[7] M. D. DONSKER, "An invariance principle for certain probability limit theorems," *Mem. Am. Math. Soc.*, No. 6, 1951, pp. 1-12.

A LIMIT THEOREM FOR THE PERIODOGRAM

By Salomon Bochner[1] and Tatsuo Kawata[2]

1. Introduction. Let $\mathcal{E}(t)$ be a real stationary process in the wide sense with mean 0 and let its covariance function and spectral function be $\rho(u)$, $F(x)$ respectively. We assume that $F(x)$ is absolutely continuous and has a spectral density function $p(x)$. The second-named author, [1], has discussed the periodogram

$$J(T) = \frac{1}{4\pi T}\left|\int_{-T}^{T} \mathcal{E}(t)e^{-i\xi t}\,dt\right|^2, \tag{1.1}$$

in case $\mathcal{E}(t)$ is stationary even of the fourth order, so that the expectation

$$E\mathcal{E}(t)\mathcal{E}(t+u)\mathcal{E}(t+v)\mathcal{E}(t+w) = P(u, v, w)$$

exists and is a function of u, v, w alone. It was also assumed that the function $Q(u, v, w)$, which is the difference between $P(u, v, w)$ and the corresponding fourth moment of a stationary Gaussian process, is the Fourier transform of a function and that the latter function satisfies the Lipschitz condition. Under these assumptions it has proven that (1.1) does not converge in mean to any random variable as $T \to \infty$, but that the covariance function of $J(T)$ and $J(T')$ does tend to a limit whenever T and T' both tend to infinity in a certain related manner, and the limiting value of the covariance function was determined.

The paper involved a rather troublesome manipulation of a Fourier integral, but we have found since that under somewhat different assumptions the complications can be reduced appreciably. In a separate publication, [2], a certain integral transformation was investigated on its own merit, and in the present paper an application of the somewhat modified approach will be made to the problem of the periodogram. The expression (1.1) will be replaced by a more general one, and as regards the difference function $Q(u, v, w)$ the assumptions will be modified as follows. We add expressly the requirement that $Q(u, v, w)$ shall be integrable in E_3, but the requirement that its Fourier transform shall satisfy the Lipschitz condition is being omitted entirely.

2. The Theorem. We shall consider the random variable

$$S(T) = \frac{1}{T}\left|\int_{-\infty}^{\infty} \mathcal{E}(t)M\left(\frac{t}{T}\right)e^{-i\xi t}\right|^2 dt \tag{2.1}$$

in place of (1.1). We shall call (2.1) a generalized periodogram of $\mathcal{E}(t)$.

Let us assume that

$$P(s_1, s_2, s_3) = Q(s_1, s_2, s_3) + P_G(s_1, s_2, s_3), \tag{2.2}$$

Received December 6, 1957.

[1] Research done in connection with U. S. Army Ordnance, Office of Ordnance Research, Contract No. D.A.—36-034—ORD-2001.

[2] Work done while the second author was a research fellow at Princeton University supported by the Rockefeller Foundation.

where

(2.3) $P_G(s_1, s_2, s_3) = \rho(s_1)\,\rho(s_2 - s_3) + \rho(s_2)\rho(s_3 - s_1) + \rho(s_3)\rho(s_1 - s_2).$

If $\mathcal{E}(t)$ is a stationary Gaussian process, then $Q(s_1, s_2, s_3) \equiv 0$. This assumption was set up first by Magness [3]; see also Parzen [4].

We assume further that

$$Q(s_1, s_2, s_3) \in L_1(E_3), \tag{2.4}$$

and that the Fourier transform of $Q(s_1, s_2, s_3)$ is also in $L_1(E_3)$, so that

$$q(x_1, x_2, x_3) = \int_{E_3} e^{i(s\cdot x)} Q(s_1, s_2, s_3)\, dv_s, \tag{2.5}$$

$$Q(s_1, s_2, s_3) = (2\pi)^{-3} \int_{E_3} e^{-i(s\cdot x)} q(x_1, x_2, x_3)\, dv_x, \tag{2.6}$$

where E_k denotes the whole Euclidean space of k dimension and $(s \cdot x) = s_1x_1 + s_2x_2 + s_3x_3$.

Under these conditions, we obtain the following theorem.

THEOREM. *Let $M(\alpha)$ be bounded and integrable in $(-\infty, \infty)$ and let the Fourier transform $K(x)$ of $M(\alpha)$*

$$K(x) = \int e^{ix\alpha} M(\alpha)\, d\alpha$$

satisfy

$$K(x) = O(|x|^{-1}), \text{ as } x \to \infty. \tag{2.7}$$

Then we have, as T_1 and T_2 tend to infinity such that $T_1/T_2 \to \mu$, $\mu \neq 0$,

$$\lim \operatorname{cov}\{S(T_1), S(T_2)\} = \begin{cases} (2\pi)^2\, (|C_\mu^{(1)}|^2 + |C_\mu^{(2)}|^2)p^2(0), & \text{if } \xi = 0, \\ (2\pi)^2\, |C_\mu^{(2)}|^2 p^2(\xi), & \text{if } \xi \neq 0, \end{cases} \tag{2.8}$$

and

$$\lim E\{S(T_1) - S(T_2)\}^2 = \begin{cases} 2(2\pi)^2\, (|C_1^{(1)}|^2 + |C_1^{(2)}|^2 - |C_\mu^{(1)}|^2 \\ \quad - |C_\mu^{(2)}|^2)p^2(0), & \text{if } \xi = 0, \\ 2(2\pi)^2\, (|C_1^{(2)}|^2 - |C_\mu^{(2)}|^2)p^2(\xi), & \text{if } (\xi) \neq 0, \end{cases} \tag{2.9}$$

provided that $p(x)$ is continuous at ξ, and the constants $C_\mu^{(j)}$ $(j = 1, 2)$ are given by

$$C_\mu^{(1)} = \mu^{\frac{1}{2}} \int_{-\infty}^{\infty} M(\alpha) M(\mu\alpha)\, d\alpha,$$

$$C_\mu^{(2)} = \mu^{\frac{1}{2}} \int_{-\infty}^{\infty} M(\alpha) \bar{M}(\mu\alpha)\, d\alpha.$$

We add a remark. If $\mu \to \infty$, or $\mu \to 0$, then $C_\mu^{(1)}$, $C_\mu^{(2)}$ converge to 0. This is easily seen from the fact that $C_\mu^{(1)} = C_{1/\mu}^{(1)}$, $C_\mu^{(2)} = \bar{C}_{1/\mu}^{(2)}$, and $|C_\mu^{(j)}| \leq \mu^{1/2} M \int_{-\infty}^{\infty} |M(\alpha)|\, d\alpha \to 0$, $(\mu \to 0)$, M being an upper bound of $M(\alpha)$.

We also note that the theorem implies that the constant

$$|C_1^{(1)}|^2 + |C_1^{(2)}|^2 - |C_\mu^{(1)}|^2 - |C_\mu^{(2)}|^2$$

must be non-negative. This can also be established directly by verifying that it is the value of the double integral

$$\frac{1}{2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [\,|A(\alpha, \beta)|^2 + (A(\alpha, \beta))^2]\, d\alpha\, d\beta,$$

where

$$A(\alpha, \beta) = M(\alpha)\bar{M}(\beta) - \mu M(\mu\alpha)\bar{M}(\mu\beta).$$

For the proof of the theorem, we first of all state as a lemma, a theorem given in [2].

LEMMA 1. *Let $M_j(\alpha)$ $(j = 0, 1, \cdots, k)$ be bounded and integrable over $(-\infty, \infty)$ and let their Fourier transforms be*

$$K_j(x) = \int_{-\infty}^{\infty} e^{i\alpha x} M_j(\alpha)\, d\alpha, \qquad (j = 0, 1, \cdots, k).$$

Put

$$K(x_1, x_2, \cdots, x_k; T_0, T_1, \cdots, T_k)$$
$$= \prod_{j=0}^{k} T_j \cdot K(T_0(x_1 + x_2 + \cdots + x_k)) \prod_{j=1}^{k} K_j(T_j x_j)$$

for any positive numbers $T_0, T_1, \cdots, T_k$. Then we have

$$\lim_{T_i\to\infty} \frac{1}{T_0}\int_{E_k} f(x_1, x_2, \cdots, x_k) K(x_1, x_2, \cdots, x_k; T_0, T_1, \cdots, T_k)\, dv_x$$
$$= C_k(2\pi)^k f(0, \cdots, 0), \qquad (i = 0, 1, \cdots, k),$$

if T_j go to infinity such that $T_0/T_j \to \mu_j$ and $\mu_j \neq 0$ $(j = 1, 2, \cdots, k)$ and $f(x_1, \cdots, x_k)$ satisfies the conditions that the function $f(x_1, \cdots, x_k)$ is continuous and belongs to $L_1(E_k)$ and its Fourier transform

$$g(\alpha_1, \alpha_2, \cdots, \alpha_k) = \int_{E_k} e^{i(\alpha \cdot x)} f(x_1, \cdots, x_k)\, dv_x$$

likewise belongs to $L_1(E_k)$. C_k is

$$C_k = \int_{-\infty}^{\infty} M_0(\alpha) \prod_{j=1}^{k} M_j(-\mu_j\, \alpha)\, d\alpha.$$

3. A lemma. For the proof of the theorem, we need one more lemma.

LEMMA 2. *Let $K_j(x)$, $(j = 1, 2)$ be a bounded function which is the Fourier transform of a bounded and integrable function $M_j(\alpha)$*

$$K_j(x) = \int_{-\infty}^{\infty} M_j(\alpha) e^{ix\alpha}\, d\alpha, \qquad j = 1, 2, \tag{3.1}$$

and let us assume that

$$K_j(x) = O(|x|^{-1}), \qquad \text{as } x \to \infty, \qquad j = 1, 2. \tag{3.2}$$

(i) If $p(x) \in L_1(-\infty, \infty)$ and continuous at $-\xi$, then

$$(T_1 T_2)^{\frac{1}{2}} \int_{-\infty}^{\infty} K_1(T_1(x+\xi))K_2(T_2(x+\xi))p(x)\,dx \tag{3.3}$$

converges to

$$2\pi \cdot C_\mu \cdot p(-\xi),$$

when $T_1, T_2 \to \infty$ *and* $T_1/T_2 \to \mu$ *and* $\mu \neq 0$, *where*

$$C_\mu = \mu^{\frac{1}{2}} \int_{-\infty}^{\infty} M_1(\beta)M_2(\mu\beta)\,d\beta.$$

(ii) *If* $p(x) \in L_1(-\infty, \infty)$, *and* $p(x)$ *continuous as* $-\xi_1$ *then*

$$(T_1 T_2)^{\frac{1}{2}} \int_{-\infty}^{\infty} K_1[T_1(x+\xi_1)]K_2[T_2(x+\xi_2)]p(x)\,dx \tag{3.4}$$

converges to zero when $T_1, T_2 \to \infty$ *such that* $T_1/T_2 \to \mu$ *and* $\mu \neq 0$ *and* $\xi_1 \neq \xi_2$.

PROOF. (i) We consider the integral

$$(T_1 T_2)^{1/2} \int_{-\infty}^{\infty} K_1(T_1 x)K_2(T_2 x)\,dx, \tag{3.5}$$

which is absolutely convergent because K_1, K_2 are bounded and satisfy (3.2). By the Parseval theorem, since $K_j(x) \in L_2(-\infty, \infty)$, we have

$$\begin{aligned}(T_1 T_2)^{1/2} \int_{-\infty}^{\infty} K_1(T_1 x)K_2(T_2 x)\,dx &= \frac{2\pi}{(T_1 T_2)^{1/2}} \int_{-\infty}^{\infty} M_1\left(\frac{\alpha}{T_1}\right) M_2\left(\frac{-\alpha}{T_2}\right) d\alpha \\ &= 2\pi(T_1/T_2)^{1/2} \int_{-\infty}^{\infty} M_1(\beta)M_2\left(-\frac{T_1}{T_2}\beta\right) d\beta.\end{aligned}$$

This converges to

$$2\pi C_\mu = 2\pi\mu^{1/2} \int_{-\infty}^{\infty} M_1(\beta)M_2(-\mu\beta)\,d\beta,$$

as is easily seen from the fact that

$$\int_{-\infty}^{\infty} |M_2(a\beta) - M_2(a_0\beta)|\,d\beta \to 0,$$

if $a \to a_0$ and $a_0 \neq 0$.

Hence it suffices to show that

$$I = (T_1 T_2)^{1/2} \int_{-\infty}^{\infty} K_1(T_1 x)K_2(T_2 x)\{p(x-\xi) - p(-\xi)\}\,dx \tag{3.6}$$

converges to zero.

We divide I into two parts:

$$I = \int_{|x|<\delta} + \int_{|x|>\delta} = I_1 + I_2,$$

where δ is taken so that $|p(x-\xi) - p(-\xi)| < \epsilon$, for $|x| < \delta$, ϵ being any assigned positive number.

We have

$$\begin{aligned} |I_1| &\leqq \epsilon(T_1T_2)^{1/2} \int_{|x|<\delta} |K_1(T_1x)K_2(T_2x)|\,dx \\ &\leqq \epsilon(T_2/T_1)^{1/2} \int_{-\infty}^{\infty} \left| K_1(u)K_2\left(\frac{T_2}{T_1}u\right)\right| du \qquad (3.7) \\ &\leqq \epsilon C \int_{-\infty}^{\infty} \frac{du}{1+u^2}, \end{aligned}$$

for some constant C, as follows from (3.2).

Next we have

$$\begin{aligned} |I_2| \leqq{}& (T_1T_2)^{1/2} \int_{|x|>\delta} |K_1(T_1x)K_2(T_2x)|\,|p(x-\xi)\,dx \\ &+ |p(\xi)|(T_1T_2)^{1/2} \int_{|x|>\delta} |K_1(T_1x)K_2(T_2x)|\,dx \\ \leqq{}& \frac{C}{(T_1T_2)^{1/2}} \int_{|x|>\delta} \frac{|p(x-\xi)|}{x^2}\,dx + \frac{C|p(\xi)|}{(T_1T_2)^{1/2}} \int_{|x|>\delta} \frac{dx}{x^2}, \end{aligned}$$

for some constant C. Hence we get

$$I_2 = o(1) \qquad (3.8)$$

as $T_1T_2 \to \infty$, and this together with (3.7) proves (i).

We shall now prove (ii). We have

$$\begin{aligned} (T_1T_2)^{1/2} &\int_{-\infty}^{\infty} K_1(T_1(x+\xi_1))K_2(T_2(x+\xi_2))\,dx \\ &= \frac{2\pi}{(T_1T_2)^{1/2}} \int_{-\infty}^{\infty} M_1\left(\frac{\alpha}{T_1}\right) M_2\left(\frac{-\alpha}{T_2}\right) e^{i\alpha(\xi_1-\xi_2)}\,d\alpha \qquad (3.9) \\ &= 2\pi(T_1/T_2)^{1/2} \int_{-\infty}^{\infty} M_1(\beta)M_2\left(-\frac{T_2}{T_1}\beta\right) e^{iT_1\beta(\xi_1-\xi_2)}\,d\beta \end{aligned}$$

and the difference between this and the expression

$$2\pi(T_1/T_2)^{1/2} \int_{-\infty}^{\infty} M_1(\beta)M_2(-\mu\beta)e^{iT_1\beta(\xi_1-\xi_2)}\,d\beta \qquad (3.10)$$

is in absolute value

$$\begin{aligned} &\leqq 2\pi(T_1/T_2)^{1/2} \int_{-\infty}^{\infty} |M_1(\beta)| \left| M_2\left(-\frac{T_1}{T_2}\beta\right) - M_2(-\mu\beta)\right| d\beta \\ &\leqq C \int_{-\infty}^{\infty} \left| M_2\left(-\frac{T_1}{T_2}\beta\right) - M_2(-\mu\beta)\right| d\beta. \end{aligned}$$

But this is as small as we please, for T_1, T_2 large and T_1/T_2 near to μ, provided $\mu \neq 0$.

Now (3.10) tends to zero by Riemann-Lebesgue lemma, and we conclude that (3.9) tends to zero also.

It suffices, then, to show that

$$J = (T_1 T_2)^{1/2} \int_{-\infty}^{\infty} K_1[T_1(x + \xi_1)]K_2[T_2(x + \xi_2)]\{p(x) - p(-\xi_1)\}\, dx \tag{3.11}$$

converges to zero.

We have

$$\begin{aligned} J &= (T_1 T_2)^{1/2} \int_{-\infty}^{\infty} K_1(T_1 y)K_2\{T_2[y - (\xi_1 - \xi_2)]\}\{p(y - \xi_1) - p(-\xi_1)\}\, dy \\ &= (T_1 T_2)^{1/2} \int_{|y|<\delta} + (T_1 T_2)^{1/2} \int_{|y|>\delta} = J_1 + J_2, \end{aligned} \tag{3.12}$$

say. Here δ is so chosen that

$$|p(y - \xi_1) - p(-\xi_1)| < \epsilon, \tag{3.13}$$

for $|y| < \delta$ and

$$|\xi_1 - \xi_2| - \delta > c > 0, \tag{3.14}$$

for some positive constant c. Then

$$\begin{aligned} |J_1| &\leqq (T_1 T_2)^{1/2} \cdot \epsilon \int_{|y|<\delta} |K_1(T_1 y)K_2[T_2 y - T_2(\xi_1 - \xi_2)]|\, dy \\ &\leqq \epsilon (T_1 T_2)^{1/2} C \int_{|y|<\delta} \frac{dy}{T_2(|\xi_1 - \xi_2| - y)} \\ &\leqq \epsilon (T_1/T_2)^{1/2} C \cdot c \cdot \delta \leqq C\epsilon, \end{aligned} \tag{3.15}$$

for some constant C by (3.13) and (3.14).

Next we shall consider J_2. We divide J_2 further into two parts,

$$\begin{aligned} J_2 &= (T_1 T_2)^{1/2} \int_{|y|>\delta,\, |y-(\xi_1-\xi_2)|>\eta} + (T_1 T_2)^{1/2} \int_{|y|>\delta,\, |y-(\xi_1-\xi_2)|<\eta} \\ &= J_{21} + J_{22}, \end{aligned}$$

say, where $0 < \eta < \frac{1}{2}|\xi_1 - \xi_2|$. Then

$$\begin{aligned} |J_{21}| &\leqq (T_1 T_2)^{1/2} C \int_{|y|>\delta,\, |y-(\xi_1-\xi_2)|>\eta} \frac{1}{T_1 y} \\ &\qquad \cdot \frac{1}{T_2 |y - (\xi_1 - \xi_2)|} (|p(y - \xi_1)| + |p(-\xi_1)|)\, dy \\ &\leqq \frac{C}{(T_1 T_2)^{1/2}\delta\eta} \int_{|y|>\delta,\, |y-(\xi_1-\xi_2)|>\eta} \frac{|p(y - \xi_1)| + |p(-\xi_1)|}{y\,|y - (\xi_1 - \xi_2)|}\, dy, \end{aligned} \tag{3.16}$$

which converges to zero as $T_1, T_2 \to \infty$, since the integral is finite. Moreover

$$
\begin{aligned}
|J_{22}| &\leqq (T_1 T_2)^{1/2} C \int_{(\xi_1-\xi_2)-\eta<y<(\xi_1-\xi_2)+\eta} \\
&\qquad \cdot \frac{1}{T_1 y} (|p(y-\xi_1)| + |p(-\xi_1)|)\, dy \\
&\leqq (T_2/T_1)^{1/2} \frac{2C}{|\xi_1-\xi_2|} \int_{(\xi_1-\xi_2)-\eta}^{(\xi_1-\xi_2)+\eta} (|p(y-\xi_1)| + |p(-\xi_1)|)\, dy \\
&\leqq C \int_{(\xi_1-\xi_2)-\eta}^{(\xi_1-\xi_2)+\eta} (|p(y-\xi_1)| + |p(-\xi_1)|)\, dy.
\end{aligned} \tag{3.17}
$$

Hence $\limsup_{T_1, T_2\to\infty, T_1/T_2\to\mu}$ of (3.17) is small for η small, that is

$$\lim J_{22} = 0. \tag{3.18}$$

From (3.16), (3.18) we obtain

$$\lim J_2 = 0,$$

which together with (3.15) gives $\lim J = 0$.

4. Proof of the theorem. We now proceed to prove the theorem stated in Section 2.

We start with the computation of

$$
\begin{aligned}
ES(T_1)S(T_2) &= \frac{1}{T_1 T_2} E \left| \int_{-\infty}^{\infty} \varepsilon(t) M\left(\frac{t}{T_1}\right) e^{-i\xi t}\, dt \right|^2 \cdot \left| \int_{-\infty}^{\infty} \varepsilon(t) M\left(\frac{t}{T_2}\right) e^{-i\xi t}\, dt \right|^2 \\
&= \frac{1}{T_1 T_2} E \int_{E_4} \varepsilon(t_1)\varepsilon(t_2)\varepsilon(t_3)\varepsilon(t_4) e^{-i\xi(t_1-t_2+t_3-t_4)} \\
&\qquad \cdot M\left(\frac{t_1}{T_1}\right) \bar{M}\left(\frac{t_2}{T_1}\right) M\left(\frac{t_3}{T_2}\right) \bar{M}\left(\frac{t_4}{T_2}\right) dv_t \\
&= \frac{1}{T_1 T_2} \int_{E_4} P(t_2-t_1, t_3-t_1, t_4-t_1) e^{-i\xi(t_1-t_2+t_3-t_4)} \\
&\qquad \cdot M\left(\frac{t_1}{T_1}\right) \bar{M}\left(\frac{t_2}{T_1}\right) M\left(\frac{t_3}{T_2}\right) \bar{M}\left(\frac{t_4}{T_2}\right) dv_t \\
&= \frac{1}{T_1 T_2} \int_{E_4} Q(t_2-t_1, t_3-t_1, t_4-t_1) e^{-i\xi(t_1-t_2+t_3-t_4)} \\
&\qquad \cdot M\left(\frac{t_1}{T_1}\right) \bar{M}\left(\frac{t_2}{T_1}\right) M\left(\frac{t_3}{T_2}\right) \bar{M}\left(\frac{t_4}{T_2}\right) dv_t \\
&\quad + \frac{1}{T_1 T_2} \int_{E_4} P_G(t_2-t_1, t_3-t_1, t_4-t_1) e^{-i\xi(t_1-t_2+t_3-t_4)} \\
&\qquad \cdot M\left(\frac{t_1}{T_1}\right) \bar{M}\left(\frac{t_2}{T_1}\right) M\left(\frac{t_3}{T_2}\right) \bar{M}\left(\frac{t_4}{T_2}\right) dv_t \\
&= S_1(T_1, T_2) + S_2(T_1, T_2).
\end{aligned}
$$

Inserting (2.6) in $S_1(T_1, T_2)$, we have

$$S_1(T_1, T_2) = (2\pi)^{-3} \frac{1}{T_1 T_2} \int_{E_4} M\left(\frac{t_1}{T_1}\right) \bar{M}\left(\frac{t_2}{T_1}\right) M\left(\frac{t_3}{T_2}\right) \bar{M}\left(\frac{t_4}{T_2}\right) \cdot e^{-i\xi(t_1-t_2+t_3-t_4)}\, dv_t$$

$$\cdot \int_{E_3} q(x_1, x_2, x_3) \exp\left[i(t_2 - t_1)x_1 + i(t_3 - t_1)x_2 + i(t_4 - t_1)x_3\right] dv_x$$

$$= (2\pi)^{-3} \frac{1}{T_1 T_2} \int_{E_3} q(x_1, x_2, x_3)\, dv_x$$

$$(4.1) \qquad \cdot \int_{E_4} M\left(\frac{t_1}{T_1}\right) \bar{M}\left(\frac{t_2}{T_1}\right) M\left(\frac{t_3}{T_2}\right) \bar{M}\left(\frac{t_4}{T_2}\right)$$

$$\cdot \exp\left[i\{-t_1(x_1 + x_2 + x_3 + \xi) + t_2(x_1 + \xi) + t_3(x_2 - \xi) + t_4(x_3 + \xi)\}\right] dv_t$$

$$= (2\pi)^{-3} T_1 T_2 \int_{E_3} q(x_1, x_2, x_3) K[-T_1(x_1 + x_2 + x_3 + \xi)]$$

$$\cdot \bar{K}[-T_1(x_1 + \xi)] \cdot K[T_2(x_2 - \xi)] \bar{K}[-T_2(x_3 + \xi)]\, dv_x$$

$$= (2\pi)^{-3} T_1 T_2 \int_{E_3} q(x_1 - \xi, x_2 + \xi, x_3 - \xi)$$

$$\cdot K[-T_1(x_1 + x_2 + x_3)] \cdot \bar{K}(-T_1 x) K(T_2 x_2) \bar{K}(-T_2 x_3)\, dv_x,$$

where we denote

$$(4.2) \qquad K(x) = \int_{-\infty}^{\infty} M(\alpha) e^{ix\alpha}\, d\alpha.$$

Since $M(x)$ and $q(x_1, x_2, x_3)$ satisfy the condition of Lemma 1, we obtain that (4.1) multiplied by T_2 is convergent when $T_1/T_2 \to \mu (\mu \neq 0)$. Hence (4.1) converges to zero.

Next we shall consider $S_2(T_1, T_2)$. Inserting (2.3), we obtain

$$S_2(T_1, T_2) = \frac{1}{T_1 T_2} \int_{E_4} \{\rho(t_2 - t_1)\rho(t_3 - t_4) + \rho(t_3 - t_1)\rho(t_4 - t_2)$$

$$(4.4) \qquad + \rho(t_4 - t_1)\rho(t_2 - t_3)\} \bar{M}\left(\frac{t_1}{T_1}\right) M\left(\frac{t_2}{T_1}\right) M\left(\frac{t_3}{T_2}\right) \bar{M}\left(\frac{t_4}{T_2}\right) \cdot e^{-i\xi(t_1-t_2+t_3-t_4)} \cdot dv_t$$

$$= U_1(T_1, T_2) + U_2(T_1, T_2) + U_3(T_1, T_2),$$

say, where

$$(4.5) \qquad U_1(T_1, T_2) = \frac{1}{T_1 T_2} \int_{E_4} \rho(t_2 - t_1)\rho(t_3 - t_4) \cdot M\left(\frac{t_1}{T_1}\right) \bar{M}\left(\frac{t_2}{T_1}\right) M\left(\frac{t_3}{T_2}\right) \bar{M}\left(\frac{t_4}{T_2}\right) \cdot e^{-i\xi(t_1-t_2+t_3-t_4)}\, dv_t,$$

and U_2, U_3 are similar terms. By the assumptions of the theorem, we have $\rho(u) = \int_{-\infty}^{\infty} e^{iux}p(x)\,dx$, and, if we insert this into (4.5), we obtain

$$U_1(T_1, T_2) = \frac{1}{T_1T_2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x)p(y)\,dx\,dy$$

$$\cdot \int_{E_4} M\left(\frac{t_1}{T_1}\right)e^{-it_1(x+\xi)}\bar{M}\left(\frac{t_2}{T_1}\right)e^{it_2(x+\xi)}M\left(\frac{t_3}{T_2}\right)e^{it_3(y-\xi)}\bar{M}\left(\frac{t_4}{T_2}\right)e^{-it_4(y-\xi)}\,dv_t$$

$$= T_1T_2\int_{-\infty}^{\infty} p(x)K[-T_1(x+\xi)]\bar{K}[-T_1(x+\xi)]\,dx$$

$$\cdot \int_{-\infty}^{\infty} p(y)K[T_2(y-\xi)]\bar{K}[T_2(y-\xi)]\,dy.$$

Since $\mathfrak{L}(t)$ is real, ρ is real too, and $p(x)$ is an even function, and hence by Lemma 2, we get

$$\lim_{T_1,T_2\to\infty} U_1(T_1, T_2) = (2\pi)^2C_1^2\,p(\xi)p(-\xi) = (2\pi C_1)^2p^2(\xi), \tag{4.6}$$

where

$$C_1 = \int_{-\infty}^{\infty}\bar{M}(\beta)M(\beta)\,d\beta = \int_{-\infty}^{\infty}|M(\beta)|^2\,d\beta. \tag{4.7}$$

Quite similarly

$$U_2(T_1, T_2) = T_1T_2\int_{-\infty}^{\infty} p(x)K[-T_1(x+\xi)]K[T_2(x-\xi)]\,dx$$

$$\cdot \int_{-\infty}^{\infty} p(y)\bar{K}[T_1(y-\xi)]\bar{K}[-T_2(y+\xi)]\,dy.$$

If $\xi = 0$, then, by (3.3),

$$U_2(T_1, T_2) \to (2\pi)^2\,|C_\mu^{(1)}|^2p^2(0), \qquad (T_1, T_2 \to \infty,\ T_1/T_2 \to \mu), \tag{4.8}$$

where

$$C_\mu^{(1)} = \mu^{1/2}\int_{-\infty}^{\infty} M(\beta)M(\mu\beta)\,d\beta. \tag{4.9}$$

If $\xi \neq 0$, then (3.4) shows

$$U_2(T_1, T_2) \to 0, \quad (T_1, T_2 \to \infty,\ T_1/T_2 \to \mu). \tag{4.10}$$

Finally we have

$$U_3(T_1, T_2) = T_1T_2\int_{-\infty}^{\infty} p(x)K[-T_1(x+\xi)]\bar{K}[-T_2(x+\xi)]\,dx$$

$$\cdot \int_{-\infty}^{\infty} p(y)\bar{K}[-T_1(y+\xi)]K[-T_2(y+\xi)]\,dy,$$

and

$$U_3(T_1, T_2) \to (2\pi)^2\,|C_\mu^{(2)}|^2p^2(\xi), \qquad \text{for every } \xi, \tag{4.11}$$

where

$$C_\mu^{(2)} = \mu^{1/2} \int_{-\infty}^{\infty} M(\beta)\bar{M}(\mu\beta)\, d\beta. \tag{4.12}$$

Inserting (4.7) (4.8) (4.10) and (4.11) into (4.4), we get: If $\xi \neq 0$

$$S_2(T_1, T_2) \to (2\pi)^2 (C_1^2 + |C_\mu^{(2)}|^2)p^2(\xi) \tag{4.13}$$

as $T_1, T_2 \to \infty$, $T_1/T_2 \to \mu(\neq 0)$, and if $\xi = 0$

$$S_2(T_1, T_2) \to (2\pi)^2(C_1^2 + |C_\mu^{(1)}|^2 + |C_\mu^{(2)}|^2). \tag{4.14}$$

Hence we get

$$ES(T_1)S(T_2) \to (2\pi)^2(C_1^2 + |C_\mu^{(2)}|^2)p^2(\xi), \qquad \text{if } \xi \neq 0, \mu \neq 0, \tag{4.15}$$

$$E(T_1)S(T_2) \to (2\pi)^2(C_1^2 + |C_\mu^{(1)}|^2 + |C_\mu^{(2)}|^2)p^2(0), \qquad \text{if } \xi = 0, \mu \neq 0. \tag{4.16}$$

We also have

$$ES(T) = \frac{1}{T}\int_{E_2} E\mathfrak{L}(t_1)\mathfrak{L}(t_2)M\left(\frac{t_1}{T}\right)\bar{M}\left(\frac{t_2}{T}\right)e^{-\xi(t_1-t_2)}\, dv_t$$

$$= T\int_{-\infty}^{\infty} p(x) \mid K[T(x-\xi)] \mid^2 dx,$$

and by Lemma 2 this converges to $2\pi C_1 p(\xi)$. Thus we find that

$$\operatorname{cov}\{S(T_1), S(T_2)\} \equiv ES(T_1)S(T_2) - ES(T_1)\cdot ES(T_2)$$

converges to

$$(2\pi)^2|C_\mu^{(2)}|^2p^2(\xi), \qquad \text{if } \xi \neq 0,$$

and to

$$(2\pi)^2(|C_\mu^{(1)}|^2 + |C_\mu^{(2)}|^2)p^2(0), \qquad \text{if } \xi = 0,$$

when T_1, T_2 increase indefinitely such as $T_1/T_2 \to \mu(\mu \neq 0)$.

Especially var $S(T)$ converges to

$$(2\pi)^2|C_\mu^{(2)}|^2p^2(\xi), \qquad \text{if } \xi = 0,$$

and to

$$((2\pi)^2(|C_\mu^{(1)}|^2 + |C_\mu^{(2)}|^2)p^2(0), \qquad \text{if } \xi = 0.$$

Also, we easily find that

$$E\,|S(T_1) - S(T_2)|^2$$

converges to

$$2(2\pi)^2(|C_1^{(2)}|^2 - |C_\mu^{(2)}|^2)p^2(\xi), \qquad \text{if } \xi \neq 0$$

and to

$$2(2\pi)^2(|C_1^{(1)}|^2 + |C_1^{(2)}|^2 - |C_\mu^{(1)}|^2 - |C_\mu^{(2)}|^2)p^2(0), \qquad \text{if } \xi = 0.$$

Hence the theorem is proved.

REFERENCES

[1] S. Bochner and T. Kawata, "Special integral transformation in Euclidean space," to be published.

[2] T. Kawata, "Some convergence theorems for the stationary stochastic process," to be published.

[3] T. A. Magness, "Special response of quadratic device to non-Gaussian noise," *J. Appl. Phys.*, Vol. 25 (1954).

[4] E. Parzen, "On consistent estimates of the spectrum of a stationary process," *Ann. Math. Stat.*, Vol. 28 (1957).

PROOF OF SHANNON'S TRANSMISSION THEOREM FOR FINITE-STATE INDECOMPOSABLE CHANNELS[1]

By David Blackwell, Leo Breiman, and A. J. Thomasian

University of California, Berkeley

1. Summary. For finite-state indecomposable channels, Shannon's basic theorem, that transmission is possible at any rate less than channel capacity but not at any greater rate, is proved. A necessary and sufficient condition for indecomposability, from which it follows that every channel with finite memory is indecomposable, is given. An important tool is a modification, for some processes which are not quite stationary, of theorems of McMillan and Breiman on probabilities of long sequences in ergodic processes.

2. Notation, definitions. For any positive integer N, we denote by $I(N)$ the set of integers $1, 2, \cdots, N$ and for any set S we denote by $S^{(N)}$ the set of N-tuples $(s_1, \cdots, s_N)$ with $s_i \varepsilon S$, $i \varepsilon I(N)$.

Let A be a fixed positive integer. A *source* is a pair (M, ϕ), where M is a finite, say $D \times D$, indecomposable Markov matrix and ϕ is a function from $I(D)$ to $I(A)$. A *channel* is a sequence of A Markov matrices $C(1), \cdots, C(A)$ of the same size, say $R \times R$, and a function ψ from $I(R)$ to $I(B)$, where B is some positive integer.

The elements of $I(D)$ and $I(R)$ will be considered as states of the source and channel respectively. The source will be considered as driving the channel as follows. If $d \varepsilon I(D)$, $r \varepsilon I(R)$ are the states of the source and channel at the beginning of a cycle, the source moves from d to a state $e \varepsilon I(D)$, selected according to the Markov transition matrix M, so that $M(d, e)$ is the probability that the new state is e, given that the initial state is d. The source then emits the number $\phi(e) \varepsilon I(A)$, which is fed into the channel. The channel then moves into a state $s \varepsilon I(R)$, selected according to the matrix $C(\phi(e))$, and emits the number $\psi(s)$, completing the cycle. A new cycle then begins, with e, s as the initial states of the source and channel. The joint motion of the source and channel is thus described by the *source-channel matrix*, which is a $DR \times DR$ Markov matrix L, with elements $L((d, r), (e, s)) = M(d, e)C(\phi(e), r, s)$. A channel will be called *indecomposable* if for every source the source-channel matrix L is indecomposable. Thus, for any source and any indecomposable channel, there is a sequence of random variables $\{(d_n, r_n), -\infty < n < \infty\}$, which is an ergodic Markov process with transition matrix L. Moreover, the joint distribution of $\{(d_n, r_n)\}$ depends only on L. McMillan [4], extending the work of Shannon [5], has shown that associated with any stationary ergodic process $\{z_k\}$ with a finite set F of

Received January 6, 1958; revised May 22, 1958.

[1] This research was supported in part by the Office of Naval Research under Contract Nonr-222(53) and in part by a research grant (No. G-3666) from the National Institutes of Health, Public Health Service.

states, is a number h, called the *entropy* of the process, such that for large N it is practically certain that the sequence of states of length N which occurs is one whose probability is about 2^{-Nh}; more precisely, for any sequence $f \varepsilon F^{(N)}$ let

$$Q_N(f) = \text{Prob } \{(z_1, \cdots, z_N) = f\}.$$

Then

$$N^{-1} \log Q_N(z_1, \cdots, z_N) \to -h \text{ in } L_1 \text{ as } N \to \infty, \tag{1}$$

where the log above and throughout this paper has base 2. Breiman [1] has shown that convergence with probability 1 also occurs in (1). For the ergodic process $\{(d_k, r_k)\}$, the processes $\{x_k = \phi(d_k)\}$, $\{y_k = \psi(r_k)\}$, $\{(x_k, y_k)\}$ are of course also ergodic; we denote their entropies by $H(X)$, $H(Y)$, $H(X, Y)$ respectively.

For a fixed indecomposable channel, the upper bound H over all sources of the number $H(X) + H(Y) - H(X, Y)$ is called, following Shannon, the *capacity* of the channel. Shannon [5] and, subsequently, McMillan [4], Feinstein [2], Hincin [3], and Wolfowitz [6] have shown that, under various hypotheses on the channel, it is possible to transmit over the channel at any rate less than its capacity, but not at any rate greater than its capacity. For a channel as defined above, this means, as in [6], the following. For a given channel, to say that it is possible to transmit at rate G means that for every $\epsilon > 0$ there is an N_0 such that for any $N \geqq N_0$ there are $2^{NG} = J$ distinct sequences $u_1, \cdots, u_J$, where each $u_j \varepsilon I(A)^{(N)}$, and J disjoint subsets $E_1, \cdots, E_J$ of $I(B)^N$ such that

$$Q(r, u_j, E_j) > 1 - \epsilon \text{ for all } j \text{ and all } r \varepsilon I(R), \tag{2}$$

where for any $r \varepsilon I(R)$, $u = (u(1), \cdots, u(N)) \varepsilon I(A)^{(N)}$, $E \subset I(B)^{(N)}$

$$Q(r, u, E) = \sum C(u(1), r, r_1) \cdots C(u(N), r_{N-1}, r_N),$$

where the sum is over those sequences $(r_1, \cdots, r_N)$ for which

$$(\psi(r_1), \cdots, \psi(r_N)) \varepsilon E.$$

Thus $Q(r, u, E)$ is the probability that the output sequence from the channel is an element of E, when the channel is initially in state r and u is the input sequence.

For a given channel, denote by H^* the upper bound of the rates G at which it is possible to transmit. We shall show that, for indecomposable channels of the type considered here, $H^* = H$, that is, it is possible to transmit at any rate less than the channel capacity, but not at a rate greater than channel capacity. Shannon and McMillan seem to have regarded $H^* \leqq H$ as more or less obvious, and devoted most of their attention to showing, under certain hypotheses, that $H \leqq H^*$. The other writers have given some attention to the inequality $H^* \leqq H$. In particular, Wolfowitz [6] obtained $H^* \leqq H$ for channels of zero memory. Our result, that $H^* = H$ for indecomposable channels, extends those obtained previously.

3. A necessary and sufficient condition for indecomposability. To verify that the results to be proved in Sections 5 and 6 are valid for a given channel, we must show that the channel is indecomposable. The following criterion is helpful.

THEOREM 1. *A channel* $(C(1), \cdots, C(A))$ *is indecomposable if and only if every finite product* $C(a_1) \cdots C(a_k)$ *is an indecomposable Markov matrix,* $k = 1, 2, \cdots,$ $a_i \in I(A)$.

PROOF. Suppose the channel is indecomposable and let $a_1, \cdots, a_k$ be any finite sequence of elements of $I(A)$. Consider the source with k states $1, \cdots, k$ with $M(i, i + 1) = 1$ for $i < k$, $M(k, 1) = 1$, and $\phi(i) = a_i$. Let

$$F = C(a_1) \cdots C(a_k)$$

and let $r_1, r_2 \in I(R)$. To show that F is indecomposable it is sufficient to find integers T_1, T_2 and a state $r_3 \in I(R)$ such that $F^{T_1}(r_1, r_3) > 0$ and $F^{T_2}(r_2, r_3) > 0$, that is, such that r_3 is reachable from either r_1 or r_2 under transition matrix F. Since the source-channel matrix L is indecomposable, the two states (k, r_1), (k, r_2) have a common possible successor (i, r) which itself has a possible successor of the form (k, r_3). Thus (k, r_3) is a possible successor of either (k, r_1) or (k, r_2). Since the source has period k, the times after which (k, r_3) can be reached from (k, r_1) or (k, r_2) are multiples of k, that is, there are integers T_1, T_2 such that $L^{T_ik}((k, r_i), (k, r_3)) > 0$ for $i = 1, 2$. But $L^{Tk}((k, r), (k, s)) = F^T(r, s)$. Consequently $F^{T_i}(r_i, r_3) > 0$ for $i = 1, 2$ and F is indecomposable.

Now suppose that every finite product $C(a_1) \cdots C(a_k)$ is indecomposable, and let (M, ϕ) be any source. Let (d, r), (e, s) be any two source-channel states; we must find a common possible successor (f, t). Since M is indecomposable, d and e have a common possible successor f which is recurrent. There are then numbers r', s', such that (f, r') is a successor of (d, r) and (f, s') is a successor of (e, s), so that any common successor of (f, r') and (f, s') is also a common successor of (d, r) and (e, s). Thus we may suppose $d = e = f$, and must find a common successor (f, t) of (f, r'), (f, s'), where f is recurrent. Let $f_0 = f, f_1, \cdots, f_{k-1}, f_k = f$ be a possible path from f to itself, and let $F = C(\phi(f_1)) \cdots C(\phi(f_k))$. We assert that if t is a possible successor of r' with respect to F, then (f, t) is a possible successor of (f, r') in the source-channel matrix L. For $L^{Tk}((f, r'), (f, t)) \geq [M(f_0, f_1) \cdots M(f_{k-1}, f_k)]^T F^T(r', t)$, and since the first factor on the right is positive, the left side is positive whenever $F^T(r', t)$ is. But since F is recurrent, r' and s' have a common possible successor t with respect to F, so that (f, t) is a common possible successor of (f, r'), (f, s) in L, completing the proof.

We shall say that a channel has memory m if every product $C(a_0) \cdots C(a_m)$ has identical rows. Thus a channel has memory m if and only if the conditional distribution of the present state of the channel, given the present input a_m, the m previous inputs $a_0, \cdots, a_{m-1}$ and the state r of the channel just prior to input a_0, is independent of r for every $a_0, \cdots, a_m$. A channel is said to have finite memory if for some m it has memory m. Every channel with finite memory is clearly indecomposable, for if $F = C(a_1) \cdots C(a_k)$, some power of F has

identical rows so that F is indecomposable. From Theorem 1, the channel is then indecomposable. That this includes, as a special case, the finite memory channels as defined by Feinstein [2] and Wolfowitz [6] can be seen from the following considerations: let the inputs to a channel be denoted by $\cdots, X_{-1}, X_0, X_1, \cdots$ and the outputs by $\cdots, Y_{-1}, Y_0, Y_1, \cdots$ and let the probability structure at the channel be defined, following McMillan [4], by specifying the conditional probabilities of the various output messages, given the input signals. That is, we are given the conditional probabilities $p(Y_n, \cdots, Y_k \mid X_n, X_{n-1}, \cdots)$ where we are now assuming that the channel is nonanticipatory and stationary. We assume, in addition, that there is an integer m such that

$$p(Y_n \mid X_n, Y_{n-1}, X_{n-1}, Y_{n-2}, X_{n-2}, \cdots) = p(Y_n \mid X_n, Y_{n-1}, X_{n-1}, \cdots, Y_{n-m}, X_{n-m}).$$

Now if we consider the finite state channel whose states consist of m-tuples of pairs, one member of the pairs being from the input alphabet and the other from the output alphabet, then the above assumption implies that this finite state channel is finitary in the sense described above, that is, it has the required Markov property. If we add the additional restriction that there is an integer M such that if two output messages m long, say y_1, y_2, are separated by a distance M, that

$$p(y_1, y_2 \mid \cdots X_1, X_0, X_{-1}, \cdots) = p(y_1 \mid \cdots, X_1, X_0, X_{-1}, \cdots)p(Y_2 \mid \cdots X_1, X_0, X_{-1})$$

then this finite state channel has finite memory M.

4. A modification of McMillan's theorem. In proving our main result, we shall need the following extension of a special case of McMillan's theorem.

THEOREM 2. *Let $d_1, d_2, \cdots$ be a Markov process with finite indecomposable transition matrix M, say $D \times D$, let ϕ be a function from $I(D)$ to $I(A)$, and let $y_n = \phi(d_n)$. For any sequence $s \in I(A)^{(N)}$ let $p(s) = P\{(y_1, \cdots, y_N) = s\}$, and let $z_N = p(y_1, \cdots, y_N)$. There is a constant h, depending only on M and ϕ, such that*

$$N^{-1} \log z_N \to -h \tag{3}$$

in L_1 and with probability 1 as $N \to \infty$.

PROOF. If the distribution of d_1 is the (unique) stationary distribution for M, the $\{y_n\}$ process is ergodic, and the theorems of McMillan [4] and Breiman [1] yield (3), with h as the entropy of the process.

For any $d \in I(D)$ and any event E, write $P_d(E)$ for $P(E \mid d_1 = d)$. Let $\lambda = (\lambda_1, \cdots, \lambda_D)$ be the stationary distribution for M, and let $Q(E) = \sum \lambda_d P_d(E)$. The theorems of McMillan and Breiman assert that

$$\frac{\log \sum_d \lambda_d z_{dN}}{N} \to -h \text{ a.e. and } L_1(Q), \tag{4}$$

where $p_d(s) = P_d\{(y_1, \cdots, y_N) = s\}$ and $z_{dN} = p_d(y_1, \cdots, y_N)$. For any d for which $\lambda_d > 0$, we have

$$\lambda_d z_{dN} = \left(\sum_e \lambda_e z_{eN}\right) Q(d_1 = d \mid y_1, \cdots, y_N).$$

Taking logs, dividing by N, letting $N \to \infty$ and using (4) and the fact that $Q(d_1 = d \mid y_1, \cdots, y_N)$ converges a.e. (Q) to a limit which is positive a.e. (P_d) yields

$$\frac{\log z_{dN}}{N} \to -h \text{ a.e. } P_d \text{ for } \lambda_d > 0. \tag{5}$$

Now let d be a state for which $\lambda_d = 0$, let e be any state for which $\lambda_e > 0$, let k be any integer $\leqq N$ and let G denote the event $\{d_k = e\}$. We have

$$z_{dN} P_d(G \mid y_1, \cdots, y_N) = z_{dk} P_d(G \mid y_1, \cdots, y_k) p_e(y_k, \cdots, y_N). \tag{6}$$

Since the P_d conditional distribution of $y_k, y_{k+1}, \cdots$, given that G occurs, is the same as the unconditional P_e distribution of $y_1, y_2, \cdots$, we conclude from (5) that on G, a.e. P_d, $N^{-1} \log p_e(y_k, \cdots, y_N) \to -h$. Also, on G, a.e. P_d, $P_d(G \mid y_1, \cdots, y_N)$ has a positive limit as $N \to \infty$ and $z_{dk} P_d(G \mid y_1, \cdots, y_k)$ is positive. Taking logs in (6), dividing by N, letting $N \to \infty$ yields

$$N^{-1} \log z_{dN} \to -h \text{ a.e. } P_d \text{ on } G. \tag{7}$$

Since the union of the sets G obtained by varying k and e has P_d measure 1, we conclude

$$N^{-1} \log z_{dN} \to -h \text{ a.e. } P_d \text{ for all } d. \tag{8}$$

Next let $\mu = (\mu_1, \cdots, \mu_D)$ be any initial distribution and let $P = \sum \mu_d P_d$. For any d for which $\mu_d > 0$, we have

$$\mu_d z_{dN} = \left(\sum_d \mu_d z_{dN}\right) P(d_1 = d \mid y_1, \cdots, y_N). \tag{9}$$

Taking logs, dividing by N, letting $N \to \infty$ and using (8) yields

$$N^{-1} \log \left(\sum \mu_d z_{dN}\right) \to -h \text{ a.e. } P_d, \tag{10}$$

from which we obtain

$$N^{-1} \log \left(\sum \mu_d z_{dN}\right) \to -h \text{ a.e. } P. \tag{11}$$

Thus the probability 1 convergence in (2) is established. Finally, to obtain L_1 convergence we note, following McMillan, that the sequence $\{N^{-1} \log z_N\}$ is uniformly integrable. We have

$$J(N,k) = \int_{B_k} \mid N^{-1} \log z_N \mid dP = N^{-1} \sum p(s) \mid \log p(s) \mid \tag{12}$$

$$\leqq (k+1) 2^{-kN} A^N,$$

where B_k is the event $\{k \leqq |N^{-1} \log z_N| < k + 1\}$ and the sum is extended over those $s \in I(A)^{(N)}$ for which $k \leqq |N^{-1} \log p(s)| < k + 1$. Choose k_1 so that $2^{-k_1} A < 1$. For $k \geqq k_1$ we have

$$J(N, k) \leqq (k + 1)2^{-k}2^{k_1}. \tag{13}$$

Thus $\sum_{k_0}^{\infty} J(N, k)$ goes to zero as $k_0 \to \infty$ uniformly in N, and uniform integrability is established, completing the proof.

5. The direct half of Shannon's theorem (possibility of transmission at every rate less than capacity). We shall need the following lemma.

LEMMA. *Let p be a probability distribution on a finite product space $X \times Y$. Write $a(x) = \sum_y p(x, y)$, $b(y) = \sum_x p(x, y)$, $p(y \mid x) = p(x, y) / a(x)$. For any numbers δ, λ such that $0 < \delta \leqq \lambda < 1$, let*

$$A = \{y : b(y) > \delta\}, \qquad B = \{(x, y) : p(y \mid x) < \lambda\}.$$

For any integer M there are M points $x_1, \cdots, x_M \in X$ and M disjoint subsets $E_1, \cdots, E_M$ of Y such that

$$\sum_{y \notin E_i} p(y \mid x_i) \leqq 4M(\delta/\lambda) + 2\sum_{y \in A} b(y) + 2\sum_{(x,y) \in B} p(x, y) \tag{14}$$

for $i = 1, \cdots, M$.

PROOF. Let $X_1, \cdots, X_{2M}$ be independent random variables with distribution $a(x)$. For each $i \in I(2M)$, $y \in Y$, we define the random variable $Z(i, y) = 1$ if $p(y \mid X_i) \leqq \max_{j \neq i} p(y \mid X_j)$, $Z(i, y) = 0$ otherwise, and define

$$f_i = \sum_y p(y \mid X_i) Z(i, y).$$

Then

$$\begin{aligned} Ef_i &= \sum_x a(x) E(f_i \mid X_i = x) = \sum_{x,y} p(x, y) E(Z(i, y) \mid X_i = x) \\ &\leqq \sum_{y \in A} b(y) + \sum_{(x,y) \in B} p(x, y) + \sum{}^* p(x, y) E(Z(i, y) \mid X_i = x), \end{aligned} \tag{15}$$

where $\sum^*$ indicates summation over pairs (x, y) for which $b(y) \leqq \delta$ and $p(y \mid x) \geqq \lambda$. Now $E(Z(i, y) \mid X_i = x) = 1 - (1 - u(x, y))^{2M-1}$, where

$$u(x,y) = \sum_{v \ni p(y|v) \geqq p(y|x)} a(v).$$

For pairs (x, y) in $\sum^*$,

$$\delta \geqq b(y) = \sum_v a(v) p(y \mid v) \geqq \lambda \sum_{p(y|v) \geqq \lambda} a(v) \geqq \lambda u(x, y),$$

so that

$$E(Z(i, y) \mid X_i = x) \leqq 1 - (1 - (\delta/\lambda))^{2M-1} \leqq 2M\delta/\lambda.$$

Using this inequality in (15) yields

$$Ef_i \leqq \sum_{y \varepsilon A} b(y) + \sum_{x,y \varepsilon B} p(x,y) + 2M\delta/\lambda = \alpha. \tag{16}$$

It follows that $E(\sum_1^{2M} f_i \,/\, 2M) \leqq \alpha$. Thus there are values of $X_1, \cdots, X_{2M}$, say $x_1^*, \cdots, x_{2M}^*$, for which $\sum_1^{2M} f_i^* \,/\, 2M \leqq \alpha$, where $f_i^* = f_i(x_1^*, \cdots, x_{2M}^*)$. Since all f_i^* are $\geqq 0$, at least M of them, say $f_{i_1}^*, \cdots, f_{i_M}^*$, are $\leqq 2\alpha$. Then $2\alpha \geqq \sum' p(y \mid x_{i_j}^*)$ where the sum is over y for which

$$p(y \mid x_{i_j}^*) \leqq \max_{i \neq i_j} p(y \mid x_i^*).$$

Denoting $x_{i_j}^*$ by x_j and the set of y for which

$$p(y \mid x_{i_j}^*) > \max_{i \neq i_j} p(y \mid x_i^*)$$

by E_j yields (14), and the lemma is proved.

THEOREM 3. *For any indecomposable channel, $H^* \geqq H$, that is, it is possible to transmit at any rate less than the capacity of the channel.*

PROOF. Let (M, ϕ) be any source and let $\{(d_n, r_n), n = 0, 1, 2, \cdots\}$ be a Markov process whose transition matrix is the source-channel matrix and with d_0, r_0 having a uniform distribution on the DR states. Let $x_n = \phi(d_n)$, $y_n = \psi(r_n)$. For any $s \varepsilon I(A)^{(N)}$, $t \varepsilon I(B)^{(N)}$, $r \varepsilon I(R)$, write

$$\begin{aligned} a(s) &= P((x_1, \cdots, x_N) = s), \qquad b(t) = P((y_1, \cdots, y_N) = t), \\ Q(r, s, t) &= P((x_1, \cdots, x_N) = s, \quad (y_1, \cdots, y_N) = t, \quad r_0 = r)R/a(s), \\ p(s, t) &= P((x_1, \cdots, x_N) = s, \quad (y_1, \cdots, y_N) = t) = a(s) \textstyle\sum_r Q(r, s, t)/R. \end{aligned}$$

According to Theorem 2, as $N \to \infty$

$$\begin{aligned} &N^{-1} \log a(x_1, \cdots, x_N) \to -H(X) \\ &N^{-1} \log b(y_1, \cdots, y_N) \to -H(Y) \\ &N^{-1} \log p(x_1, \cdots, x_N, y_1, \cdots, y_N) \to -H(X, Y). \end{aligned}$$

Given $\epsilon > 0$, choose N so large that, with probability $\geqq 1 - \epsilon$,

$$\frac{\log p(x_1, \cdots, x_N, y_1, \cdots, x_N) - \log a(x_1, \cdots, x_N)}{N} \geqq H(X) - H(X,Y) - \epsilon$$

and

$$\frac{\log b(y_1, \cdots, y_N)}{N} \leqq -H(Y) + \epsilon.$$

We apply the lemma to the product space $U \times V$, where $U = I(A)^{(N)}$, $V = I(B)^{(N)}$, with $p(u, v)$ as defined above and $\delta = 2^{-N(H(Y)-\epsilon)}$, $\lambda = 2^{-N(H(X,Y)-H(X)-\epsilon)}$, and conclude the existence of $M = 2^{NG}$, say, points $u_1, \cdots, u_M \varepsilon X$ and M disjoint subsets $E_1, \cdots, E_M$ of V such that

$$\sum_{u \bar{\varepsilon} E_i} p(u \mid v_i) \leqq 4 \cdot 2^{-N[H(Y)+H(X)-H(X,Y)-G-2\epsilon]} + 8\epsilon.$$

Thus for any $G < H(X) + H(Y) - H(X, Y)$ we can, for any $\beta > 0$, by first choosing ϵ sufficiently small (less than min $(\beta/9, (H(X) + H(Y) - H(X, Y) - G)/2)$ and then choosing N sufficiently large, find $M = 2^{NG}$ X-sequences $u_1, \cdots, u_M$ of length N and M disjoint subsets $E_1, \cdots, E_M$ of $I(B)^{(N)}$ such that

$$\sum_{v \varepsilon E_i} p(v \mid u_i) > 1 - \beta. \tag{17}$$

This does not quite prove that it is possible to transmit at rate G as defined above, since (2) requires that

$$\sum_{y \varepsilon E_i} Q(r, u_i, v) > 1 - \epsilon \quad \text{for all} \quad r \varepsilon R,$$

that is, that for each initial state of the channel, each of the M messages can be correctly recovered, with large probability. This is an immediate consequence of (17), however, since (17) yields

$$R^{-1} \sum_r \left(\sum_{v \varepsilon E_i} Q(r, u_i, v) \right) > 1 - \beta,$$

so that, since $Q(r, u_i, E_i) \leqq 1$ for all r, i,

$$\sum_{v \varepsilon E_i} Q(r, u_i, v) > 1 - R\beta$$

for each r. Since β can be made arbitrarily small and R is a fixed number, the number of states of the channel, the proof is complete.

6. The converse half of Shannon's theorem (impossibility of transmission at a rate greater than capacity).

THEOREM 4. *For any indecomposable channel, $H^* \leqq H$, that is, it is not possible to transmit at a rate greater than the capacity of the channel.*

PROOF. Suppose that it is possible to transmit over a given channel at rate G, let ϵ be given, $0 < \epsilon < \frac{1}{2}$ and let $N, u_1, \cdots, u_J, J = 2^{NG}, E_1, \cdots, E_J$ denote the quantities whose existence is implied by the possibility of transmission at rate G. We may suppose that $\bigcup E_j = I(B)^{(N)}$, since if (2) is satisfied for E_j it is also satisfied if E_j is replaced by a superset. We must exhibit a source (M, ϕ) for which $H(X) + H(Y) - H(X, Y)$ is nearly G. Our source produces inputs in blocks of N by selecting one of the u_j at random, successive choices being independent. The entropy $H(X)$ will then be precisely G. Since observing a long y sequence nearly identifies the corresponding x sequence, the conditional entropy $H(X, Y) - H(Y)$ is small, so that $H(X) + H(Y) - H(X, Y)$ is nearly G.

More precisely, the input source will have NJ states (n, j), with $M((n, j), (n + 1, j)) = 1$ for $n < N$, $M((N, j), (1, i)) = 1/J$ for $i \varepsilon I(J)$. We define $\phi(n, j) = u_{jn}$, the nth symbol in the sequence u_j. Let (d_k, r_k) be a Markov process whose transition matrix is the source-channel matrix, and whose initial distribution is such that $d_1 = (1, i)$ with probability $1/J$, $i \varepsilon I(J)$ and write $x_k = \phi(d_k)$, $y_k = \psi(r_k)$. Then every x sequence of length NT which is possible has probability $J^{-T} = 2^{-NTG}$ (since $\epsilon < \frac{1}{2}$, $u_i \neq u_j$ for $i \neq j$). From Theorem 2, $H(X) = G$.

To estimate $H(X, Y) - H(Y)$, we recall some results of Shannon [5]. If x is any random variable assuming T distinct values with probabilities $p_1, \cdots, p_T$, the number $-\sum p_t \log p_t$ is called the *entropy* of x and will be denoted by $h(x)$. Always $h(x) \leqq \log T$. If (x, y) are two random variables, each with a finite set of values, the number $h(x, y) - h(y)$ is called the conditional entropy of x given y and is denoted by $h(x \mid y)$. It equals the expected value of the entropy of the conditional distribution of x given y. For any function ϕ defined on the range of y, $h(\phi(y)) \leqq h(y)$ and $h(x \mid \phi(y)) \geqq h(x \mid y)$.

Notice that, in the notation of Theorem 2, $E \log z_N = -h(y_1, \cdots, y_N)$, so that the L_1 convergence in (2) implies that $h(y_1, \cdots, y_N)/N \to H$ as $N \to \infty$. Thus, in our present notation,

$$h(x_1, \cdots, x_{NT} \mid y_1, \cdots, y_{NT})/NT \to H(X, Y) - H(Y)$$

as $T \to \infty$. We have

$$\begin{aligned} h(x_1, \cdots, x_{NT} \mid y_1, \cdots, y_{NT}) &\leqq \sum_{t=0}^{T-1} h(x_{Nt+1}, \cdots, x_{Nt+N} \mid y_{Nt+1}, \cdots, y_{Nt+N}) \\ &\leqq \sum_{t=1}^{T-1} h(a_t \mid b_t), \end{aligned}$$

where $a_t = (x_{Nt+1}, \cdots, x_{Nt+N})$ and $b_t = u_j$ if $(y_{Nt+1}, \cdots, y_{Nt+N}) \in E_j$ (we may suppose that $\bigcup E_j = I(B)^{(N)}$). We estimate $h(a_t \mid b_t)$ by the following lemma.

Lemma. *For any distribution α on a product space $U \times U$ of pairs (a, b) such that $\sum_j \alpha(a, a) \geqq 1 - \epsilon > \frac{1}{2}$ we have*

$$h(a \mid b) \leqq -g(\epsilon) + \epsilon \log (J - 1),$$

where $g(t) = t \log t + (1 - t) \log (1 - t)$, $0 \leqq t \leqq 1$, and J is the number of elements of U.

Proof of the Lemma. Let $\beta(b) = \sum_a \alpha(a, b)$. Then

$$-h(a \mid b) = \sum_b \beta(b) \sum_a \frac{\alpha(a, b)}{\beta(b)} \log \frac{\alpha(a, b)}{\beta(b)}.$$

Now

$$\begin{aligned} &\sum_j \frac{\alpha(a, b)}{\beta(b)} \log \frac{\alpha(a, b)}{\beta(b)} \\ &\qquad = \frac{\alpha(b, b)}{\beta(b)} \log \frac{\alpha(b, b)}{\beta(b)} + \frac{\beta(b) - \alpha(b, b)}{\beta(b)} \\ &\qquad \cdot \sum_{a \neq b} \frac{\alpha(a, b)}{\beta(b) - \alpha(b, b)} \log \frac{\alpha(a, b)}{\beta(b) - \alpha(b, b)} + \frac{\beta(b) - \alpha(b, b)}{\beta(b)} \log \frac{\beta(b) - \alpha(b, b)}{\beta(b)} \\ &\qquad = g\left(\frac{\alpha(b, b)}{\beta(b)}\right) - \frac{\beta(b) - \alpha(b, b)}{\beta(b)} \log (J - 1). \end{aligned}$$

Consequently,

$$-h(a \mid b) \geqq \sum_b \beta(b) g\left(\frac{\alpha(b, b)}{\beta(b)}\right) - \epsilon \log (J - 1).$$

Since $g(t)$ is convex and $\sum_b \beta(b) = 1$,

$$\sum_b \beta(b) g\left[\frac{\alpha(b, b)}{\beta(b)}\right] \geqq g[\sum_b \alpha(b, b)] \geqq g(1 - \epsilon) = g(\epsilon).$$

The hypotheses of the lemma are satisfied for (a_t, b_t), so that

$$h(a_t \mid b_t) \leqq -g(\epsilon) + \epsilon \log J = -g(\epsilon) + \epsilon NG.$$

Thus

$$h(x_1, \cdots, x_{NT} \mid y_1, \cdots, y_{NT}) \leqq T(-g(\epsilon) + \epsilon NG).$$

Dividing by NT and letting $T \to \infty$ yields

$$H(X, Y) - H(Y) \leqq -\frac{g(\epsilon)}{N} + \epsilon G.$$

Thus, assuming that transmission at rate G is possible we have for every $\epsilon > 0$ and arbitrarily large N, exhibited a source for which

$$H(X) + H(Y) - H(X, Y) \geqq G(1 - \epsilon) + g(\epsilon)/N.$$

It follows that $H \geqq H^*$ and the proof is complete.

7. Another form of Shannon's Theorem. Let $\{w_n, n = 1, 2, \dots\}$ be any stationary ergodic process whose variables have a finite set of values, say $I(W)$, and consider a given indecomposable channel as defined above. Shannon enquires whether the channel is adequate for transmitting the information produced by the source, with large probability of correct reception. To say that the channel is adequate means that, for every $\epsilon > 0$, there is an integer N_0 such that for any $N \geqq N_0$ there are (1) a function f (the encoder) from $I(W)^{(N)}$ to $I(A)^{(N)}$ and (2) a function g (the decoder) from $I(B)^{(N)}$ to $I(W)^N$ such that, for every initial state r of the channel,

$$\pi_r\{\alpha = \beta\} > 1 - \epsilon,$$

where α and β are random variables (the first N symbols produced by the source and the decoded estimate for these symbols respectively) whose joint distribution π_r is defined by

$$\pi_r\{\alpha = v, \beta = v'\} = \text{Prob}\,\{(w_1, \cdots, w_N) = v\} \sum_{g(\delta) = v'} Q(r, f(v), \delta)$$

where $Q(r, u, \delta)$, as defined earlier, is the probability that the channel, when initially in state r, on receiving an input u, will produce output δ. The form in which Shannon describes his result is the following.

THEOREM 5. *An indecomposable channel of capacity H is adequate for the stationary ergodic source $\{w_n\}$ if the entropy h of $\{w_n\}$ is less than H, and not if $h > H$.*

The idea of the proof of this result, based on McMillan's theorem and Theorems 3 and 4 above, is extremely simple. According to McMillan's theorem, the source w_n is very likely to produce one of about 2^{hN} sequences of length N, each of which has probability about 2^{-hN}. Accordingly, to have a large probability of transmitting the actual sequence accurately, it is necessary and sufficient that the channel be able to distinguish among about 2^{hN} different input sequences of length N which, by Theorem 3, it is if $h < H$ and is not if $h > H$. The proof below simply makes this idea precise.

PROOF. From (1), for any $\epsilon > 0$ there is an N_1 such that for any $N \geqq N_1$ there is a set $F \subset I(W)^{(N)}$ with not more than $2^{(h+\epsilon)N}$ elements such that

$$\text{Prob}\,\{(w_1, \cdots, w_N)\ \varepsilon\ F\} > 1 - \epsilon.$$

From Theorem 3 there is an $N_0 \geqq N_1$ such that for any $N \geqq N_0$ there are $2^{(H-\epsilon)N} = J$ distinct sequences $u_1, \cdots, u_J$ in $I(A)^{(N)}$ and J disjoint subsets $E_1, \cdots, E_J$ of $I(B)^{(N)}$ such that

$$Q(r, u_j, E_j) > 1 - \epsilon \text{ for all } j \text{ and } r.$$

If $H - \epsilon \geqq h + \epsilon$ there are at most J elements in F, so that there is a function f from $I(W)^{(N)}$ to $I(A)^N$ such that f maps distinct elements of F onto distinct u_j. With this f and with g chosen so that $g(\delta)\ \varepsilon\ F$, $f[g(\delta)] = u_j$ for all $\delta\ \varepsilon\ E_j$, we have

$$\pi_r\{\alpha = \beta\} > (1 - \epsilon)^2,$$

since the probability that $\alpha\ \varepsilon\ F$ is greater than $1 - \epsilon$ and the conditional probability, given that $\alpha = \alpha_0\ \varepsilon\ F$, that $\beta = \alpha_0$ is at least $Q(r, u_j, E_j) > 1 - \epsilon$, where $f(\alpha_0) = u_j$. Thus if $h < H$, the channel is adequate.

Conversely, suppose the channel is adequate. From (1), for any $\epsilon > 0$ there is an N_2 such that for any $N \geqq N_2$ there is a set $F_1 \subset I(W)^{(N)}$ such that

$$\text{Prob}\,\{(w_1, \cdots, w_N)\ \varepsilon\ F_1\} > 1 - \epsilon$$

and Prob $(w_1, \cdots, w_N) = \alpha_0 < 2^{-(h-\epsilon)N}$ for all $\alpha_0\ \varepsilon\ F_1$. Also, there is an $N_3 \geqq N_2$ such that for every $N \geqq N_3$ there are functions f, g satisfying the definition of adequacy. Since $\pi_r\{\alpha = \beta\} > 1 - \epsilon$, there is a subset F_2 of $I(W)^{(N)}$ such that $\pi_r\{\alpha\ \varepsilon\ F_2\} > 1 - \sqrt{\epsilon}$ the conditional probability

$$\pi_r\{\alpha = \beta \mid \alpha = \alpha_0\} > 1 - \sqrt{\epsilon} \text{ for } \alpha_0\ \varepsilon\ F_2\,.$$

Then $\pi_r\{\alpha\ \varepsilon\ F_1 \cap F_2\} > 1 - \epsilon - \sqrt{\epsilon}$, so that $F_1 \cap F_2$, and hence F_2 has at least $2^{(h-\epsilon)N}(1 - \epsilon - \sqrt{\epsilon}) = J_1$ elements. For $\alpha_0\ \varepsilon\ F_2$, define $E(\alpha_0)$ as the set of all $\delta\ \varepsilon\ I(B)^{(N)}$ such that $g(\delta) = \alpha_0$. The assertion $\pi_r\{\alpha = \beta \mid \alpha = \alpha_0\} > 1 - \sqrt{\epsilon}$ is equivalent to

$$Q(r, f(\alpha_0), E(\alpha_0)) > 1 - \sqrt{\epsilon}. \tag{17}$$

Note that, since the sets $E(\alpha_0)$ are disjoint, so are the elements of (α_0), provided $\epsilon < .707$, which we may assume. In summary, for every $\epsilon > 0$ we have found an

N_3 such that for any $N \geqq N_3$ there are at least $J_1(N, \epsilon)$ distinct elements of $I(A)^{(N)}$ (namely the $f(\alpha_0)$, $\alpha_0 \varepsilon F_2$ and $J_1(N, \epsilon)$ corresponding subsets of $I(B)^{(N)}$ (namely the $E(\alpha_0)$) such that (17) holds. Thus if $g < h$, it is possible to transmit at rate G, since, for sufficiently small ϵ ($< h - G$), $J_1(N, \epsilon) > 2^{NG}$ for all sufficiently large N. It now follows from Theorem 4 that $h \leqq H$, and the proof is complete.

REFERENCES

[1] L. BREIMAN, "The individual ergodic theorem of information theory," *Ann. Math. Stat.*, Vol. 28, No. 3 (1957), pp. 809-811.

[2] A. FEINSTEIN, "A new basic theorem of information theory," *Trans. IRE PGIT* (1954), pp. 2-22.

[3] A. I. KHINCHIN, *Mathematical Foundations of Information Theory*, Dover Publications, Inc., 1957.

[4] B. MCMILLAN, "The basic theorems of information theory," *Ann. Math. Stat.*, Vol. 24, No. 2 (1953), pp. 196-219.

[5] C. E. SHANNON, "A mathematical theory of communication," *Bell System Technical Journal*, Vol. 27 (1948), pp. 379-423, and 623-656.

[6] J. WOLFOWITZ, "The coding of messages subject to chance errors," *Illinois Journal of Mathematics*, Vol. 1 (1957), pp. 591-606.

ON THE LIMITING POWER FUNCTION OF THE FREQUENCY CHI-SQUARE TEST[1]

By Sujit Kumar Mitra

University of North Carolina[2]

1. Introduction. Several authors have recently investigated the power function of the frequency χ^2-test. Eisenhart [1] and Patnaik [2] have obtained large sample expressions for the power of the simple goodness of fit χ^2-test (i.e. where the class probabilities are completely specified by the null hypothesis). The more complicated case, in which the parameters occurring in the expression for class probabilities require to be estimated, has not received a unified treatment, although the problem has been treated in a number of specific situations by different authors, including, Patnaik [3], Sillito [4], Stevens [5], Pearson and Merrington [6], Poti [7], Chiang [8] and Taylor [9].

Due to difficulties in obtaining the power function of the frequency χ^2-test in the usual manner, Cochran, in an expository article [10] has suggested the derivation of its Pitman limiting power [11], and he illustrated it in the case of the simple goodness of fit test. The concept of asymptotic power suggested by Pitman has also been extensively used in various other areas like nonparametric inference (see e.g. Hoeffding and Rosenblatt [12]) and seems to be a useful tool for comparing alternative consistent tests or alternative designs for experimentation, with regard to their performance in the immediate neighbourhood of the null hypothesis.

The consistency of the frequency χ^2-test has already been established by Neyman [13]. The object of the present paper is to obtain the Pitman limiting power of this test when the unknown parameters occurring in the specification of class probabilities are estimated from the sample by an asymptotically efficient method like the method of maximum likelihood, minimum χ^2 etc. In section 5, we discuss a few applications of the Pitman limiting power for frequency χ^2-tests.

2. Pitman's concept of limiting power [11]. Let H_0 be a certain hypothesis and $\mathfrak{I}$ a test-procedure for testing H_0, which determines the critical region w_n in R_{N_n} (the sample space of N_n dimensions), for $n = 1, 2, \cdots$, ad. inf. Let us assume further that

$$N_{n+1} > N_n \qquad \text{for all } n, \tag{2.1}$$

$$0 < \lim_{n\to\infty} \text{Prob}\,\{w_n \mid H_0\} = \alpha < 1, \tag{2.2}$$

Received July 6, 1956; revised January 7, 1958.

[1] This work was supported in part of the United States Air Force through the Office of Scientific Research of the Air Research and Development Command.

[2] Present address: Indian Statistical Institute, Calcutta—35.

and for any alternative H

$$\lim_{n\to\infty} \text{Prob } \{w_n \mid H\} = 1 \tag{2.3}$$

Let $\{H_{0n}\}$ be a family of alternative hypotheses such that

$$\lim_{n\to\infty} \text{Prob } \{w_n \mid H_{0n}\} = \beta(\mathfrak{I}, \{H_{0n}\}) \tag{2.4}$$

exists and $0 < \beta(\mathfrak{I}, \{H_{0n}\}) < 1$.

We call $\beta(\mathfrak{I}, \{H_{0n}\})$ the limiting power of $\mathfrak{I}$ with respect to the family of alternatives $\{H_{0n}\}$.

This concept of limiting power derives its usefulness from the fact that, if $\mathfrak{I}'$ is any other test procedure, which suggests critical regions w_n', instead of w_n, with w_n' satisfying (2.2) and (2.3), and if

$$\beta(\mathfrak{I}, \{H_{0n}\}) \lesseqgtr \beta(\mathfrak{I}', \{H_{0n}\})$$

then for n sufficiently large

$$\text{Prob } \{w_n \mid H_{0n}\} \lesseqgtr \text{Prob } \{w_n' \mid H_{0n}\}.$$

3. A theorem in frequency chi-square. Suppose that we have $R = \sum_{i=1}^{q} r_i$ functions $p_{ij}(\alpha_1, \alpha_2, \cdots, \alpha_s)$, $(i = 1, 2, \cdots, q; j = 1, 2, \cdots, r_i)$, of $s < R - q$ parameters $\alpha_1, \alpha_2, \cdots, \alpha_s$ such that for all points of a non-degenerate interval A in the s-dimensional space of the α_k's the p_{ij} satisfy the following conditions

(a) $\displaystyle\sum_{j=1}^{r_i} p_{ij}(\alpha_1, \alpha_2, \cdots, \alpha_s) = 1$ for $i = 1, 2, \cdots, q$,

(b) $p_{ij}(\alpha_1, \alpha_2, \cdots, \alpha_s) > c^2 > 0$ for all ij,

(c) Every p_{ij} has continuous derivatives $\dfrac{\partial p_{ij}}{\partial \alpha_k}$ and $\dfrac{\partial^2 p_{ij}}{\partial \alpha_k \partial \alpha_l}$,

(d) The matrix $D = \left\{\dfrac{\partial p_{ij}}{\partial \alpha_k}\right\}_{R\times s}$ is of rank s.

(We shall assume that the index pairs (i, j), indicating the rows of the above matrix or of any such matrix we define in future, are arranged in the lexicographic order.)

For $n = 1, 2, \cdots$, ad. inf., let $(N_1^{(n)}, N_2^{(n)}, \cdots, N_q^{(n)})$ be a sequence of row vectors such that for $i = 1, 2, \cdots, q$, and every n, (i) $N_i^{(n)}$ is a natural number, (ii) $N_i^{(n+1)} > N_i^{(n)}$, (iii) if $N_n = \sum_{i=1}^{q} N_i^{(n)}$, then $N_i^{(n)}/N_n = Q_i$ independent of n.

Let $\boldsymbol{\alpha}_0' = (\alpha_1^0, \alpha_2^0, \cdots, \alpha_s^0)$ be an inner point of A and let

$$c_{ij}(i = 1, 2, \cdots, q; j = 1, 2, \cdots, r_i)$$

be a given set of numbers such that

$$\sum_{j=1}^{r_i} c_{ij} = 0, \qquad \text{for } i = 1, 2, \cdots, q. \tag{3.1}$$

Put

$$p_{ij}^0 = p_{ij}(\alpha_1^0, \alpha_2^0, \cdots, \alpha_s^0) \tag{3.2}$$

and

$$p_{ijn} = p_{ij}^0 + \frac{c_{ij}}{\sqrt{N_n}}. \tag{3.3}$$

Let n_0 be a positive integer such that for $n \geqq n_0$

$$p_{ijn} > 0 \qquad \text{for all } i, j.$$

For $n = n_0, n_0 + 1, \cdots$, ad. inf., let $\{v_{ijn}\}(i = 1, 2, \cdots, q, j = 1, 2, \cdots, r_i)$ be a sequence of R-dimensional random variables such that

$$\text{Prob } \{v_{ijn}\} = \prod_{i=1}^{q} \frac{N_i^{(n)}!}{\prod_{j=1}^{r_i} v_{ijn}!} \prod_{j=1}^{r_i} p_{ijn}^{v_{ijn}}, \tag{3.4}$$

if v_{ijn} are any set of non-negative integers (some of which might be zero) and

$$\sum_{j=1}^{r_i} v_{ijn} = N_i^{(n)}, i = 1, 2, \cdots q,$$

$$= 0, \text{ otherwise.}$$

Consider the system of equations:

$$\sum_{i=1}^{q} \sum_{j=1}^{r_i} \frac{v_{ijn} - N_n Q_i p_{ij}}{p_{ij}} \frac{\partial p_{ij}}{\partial \alpha_k} = 0, \qquad k = 1, 2, \cdots, s. \tag{3.5}$$

We shall prove

THEOREM 3.1.

(i) *The system of equations* (3.5) *have exactly one system of solutions*

$$\hat{\boldsymbol{\alpha}}_n' = (\hat{\alpha}_{1n}, \hat{\alpha}_{2n}, \cdots, \hat{\alpha}_{sn})$$

such that $\hat{\alpha}_n'$ *converges in probability to* $\hat{\boldsymbol{\alpha}}_0'$ *as* $n \to \infty$ (or, in symbols, $\hat{\boldsymbol{\alpha}}_n \xrightarrow[p]{} \boldsymbol{\alpha}_0$ as $n \to \infty$).

(ii) *The value of* χ^2 *obtained by inserting* $\alpha_k = \hat{\alpha}_{kn}$ *in*

$$\chi^2 = \sum_{i=1}^{q} \sum_{j=1}^{r_i} \frac{(v_{ijn} - N_n Q_i p_{ij}(\alpha_1, \alpha_2, \cdots, \alpha_s))^2}{N_n Q_i p_{ij}(\alpha_1, \alpha_2, \cdots, \alpha_s)} \tag{3.6}$$

is, in the limit as $n \to \infty$, distributed in a non-central χ^2-distribution ([2], [14]), with $R - s - q$ degrees of freedom and non-centrality parameter

$$\lambda = \boldsymbol{\delta}'[I - B(B'B)^{-1}B']\boldsymbol{\delta},$$

where

$$\boldsymbol{\delta} = \left\{\frac{c_{ij}\sqrt{Q_i}}{\sqrt{p_{ij}^0}}\right\}_{R\times 1},$$

and

$$B = \left\{ \frac{\sqrt{Q_i}}{\sqrt{p_{ij}^0}} \left(\frac{\partial p_{ij}}{\partial \alpha_k} \right)_{\alpha' = \alpha_0'} \right\}_{R \times s}$$

PROOF OF (i). We observe that for $\eta > d = \max_{i,j} Q_i \mid c_{ij} \mid$,

$$\mid v_{ijn} - N_n Q_i p_{ij}^0 \mid \geqq \eta\sqrt{N_n} \Rightarrow \mid v_{ijn} - N_n Q_i p_{ijn} \mid \geqq (\eta - Q_i \mid c_{ij} \mid)\sqrt{N_n} .$$

Hence, using Chebyshev's inequality, we get

$$\text{Prob} \{ \mid v_{ijn} - N_n Q_i p_{ij}^0 \mid \geqq \eta\sqrt{N_n} \} \leqq \frac{p_{ijn}(1 - p_{ijn})Q_i}{(\eta - Q_i \mid c_{ij} \mid)^2} < \frac{Q_i p_{ijn}}{(\eta - d)^2}$$

Consequently, the probability that we have $\mid v_{ijn} - N_n Q_i p_{ijn} \mid \geqq \eta\sqrt{N_n}$ for at least one subscript (i, j), is smaller than $(\eta - d)^{-2} \sum_i Q_i \sum_j p_{ijn} = (\eta - d)^{-2}$. Thus with a probability greater than $1 - (\eta - d)^{-2}$, we have

$$\mid v_{ijn} - N_n Q_i p_{ij}^0 \mid < \eta\sqrt{N_n} \qquad \text{for all } (i, j)$$

If we put

$$x_{ijn} = \frac{v_{ijn} - N_n Q_i p_{ij}^0}{\sqrt{N_n Q_i p_{ij}^0}}$$

and $a^2 = \min Q_i$, this will imply that with a probability greater than

$$1 - (\eta - d)^{-2},$$

we have

$$\mid x_{ijn} \mid < \frac{\eta}{ac} \qquad \text{for all } (i, j). \tag{3.7}$$

The proof of Theorem 3.1 (i) can now be completed using (3.7), as well as assumptions (a), (b), (c), and (d) and following Cramér's argument ([15], section 30.3).

PROOF OF (ii). We put

$$y_{ijn} = \frac{v_{ijn} - N_n Q_i p_{ij}(\hat{\alpha}_{1n}, \hat{\alpha}_{2n}, \cdots, \hat{\alpha}_{sn})}{\sqrt{N_n Q_i p_{ij}(\hat{\alpha}_{1n}, \hat{\alpha}_{2n}, \cdots, \hat{\alpha}_{sn})}}$$

$$X_{(n)} = \{x_{ijn}\}_{R \times 1}$$

$$Y_{(n)} = \{y_{ijn}\}_{R \times 1}$$

$$Z_{(n)} = \{z_{ijn}\}_{R \times 1} = Y_{(n)} - [I - B(B'B)^{-1}B']X_{(n)}$$

The proof of Theorem 3.1 (ii) requires the following results.

LEMMA 3.1.

$$z_{ijn} \xrightarrow[p]{} 0 \quad \text{as} \quad n \to \infty$$

Lemma 3.1 can be proved in a manner similar to the proof given in section 30.3 of Cramér's book [15].

LEMMA 3.2. *The limiting distribution of $X'_{(n)}$ is multivariate normal with mean. $\mathbf{\delta}'$ and covariance matrix*

$$\Lambda_X = I - PP'$$

where

$$P = \{p^0_{ij}\delta_{il}\}_{R\times q}, \qquad (i = 1, 2, \cdots, q, \; j = 1, 2, \cdots, n_i), (l = 1, 2, \cdots, q)$$

and δ_{il} is the Kronecker's symbol.

A proof of Lemma 3.2 could be constructed again, on lines similar to that in [15] section 30.1 (see also [16] p. 118.)

LEMMA 3.3. (Cramér's proposition 22.6 [15]). *Suppose that we have for $v = 1, 2, \cdots$*

$$y_v = Ax_v + z_v,$$

where x_v, y_v and z_v are n-dimensional random variables, while A is a matrix of order $n \cdot n$ with constant elements. Suppose further that, as $v \to \infty$, the distribution of x_v tends to a certain limiting distribution, while z_v converges in probability to zero. Then y_v has the limiting distribution defined by the linear transformation $y = Ax$, where x has the limiting distribution of the x_v.

LEMMA 3.4. *The limiting distribution of $Y'_{(n)}$ is multivariate normal with mean*

$$\mathbf{\delta}'[I - B(B'B)^{-1}B']$$

and covariance matrix

$$\begin{aligned}\Lambda_Y &= [I - B(B'B)^{-1}B'][I - PP'][I - B(B'B)^{-1}B'] \\ &= I - B(B'B)^{-1}B' - PP' \text{ (since } B'P = 0 \text{ as may be verified).}\end{aligned}$$

Lemma 3.4 is a direct consequence of the previous lemmas.

LEMMA 3.5. *There exists an orthogonal matrix L of order $R \cdot R$ such that*

$$L'(I - B(B'B)^{-1}B - PP')L = \begin{array}{c c} & \begin{array}{cc} s+q & R-s-q \end{array} \\ \begin{array}{c} s+q \\ R-s-q \end{array} & \begin{pmatrix} 0 & 0 \\ 0 & I \end{pmatrix} \end{array}$$

To prove Lemma 3.5, we write

$$M(R \times 2R) = [B(B'B)^{-1}B' \vdots PP']$$

and observe that

$$B(B'B)^{-1}B' + PP' = MM'$$

Since Rank $[B(B'B)^{-1}B] = s$, Rank $[PP'] = q$ and $B'P = 0$ it follows that Rank $[M] = s + q$. Hence Rank $[B(B'B)^{-1}B + PP'] = s + q$. But

$$B(B'B)^{-1}B + PP'$$

is an idempotent matrix. Hence its only nonzero latent root is 1, which is thus of multiplicity $s + q$. Therefore, since $B(B'B)^{-1}B' + PP'$ is a symmetric matrix, there exists an orthogonal matrix

$$L = [\underset{s+q}{L_1} \vdots \underset{R-s-q}{L_2}]R$$

such that

$$L'(B(B'B)^{-1}B' + PP')L = \begin{array}{c c} & \begin{array}{cc} s+q & R-s-q \end{array} \\ \begin{array}{c} s+q \\ R-s-q \end{array} & \begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix} \end{array}$$

The same matrix L satisfies Lemma 3.5.

If we now make an orthogonal transformation

$$W'_{(n)} = (w_{1,n} w_{2,n}, \cdots, w_{R,n}) = Y'_{(n)}L$$

it will then follow that the limiting distribution of $W'_{(n)}$ is multivariate normal with mean

$$\boldsymbol{\theta}' = \boldsymbol{\delta}'[I - B(B'B)^{-1}B']L$$

and covariance matrix

$$\Lambda_w = \begin{array}{c c} & \begin{array}{cc} s+q & R-s-q \end{array} \\ \begin{array}{c} s+q \\ R-s-q \end{array} & \begin{pmatrix} 0 & 0 \\ 0 & I \end{pmatrix} \end{array}$$

But

$$B(B'B)^{-1}B + PP' = [L_1\ L_2]\begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix}\begin{bmatrix} L_1' \\ L_2' \end{bmatrix} = L_1\ L_1'$$

Therefore

$$\begin{aligned} I - B(B'B)^{-1}B' &= I - L_1L_1' + PP' \\ &= L_2L_2' + PP', \quad \text{since} \quad LL' = L_1L_1' + L_2L_2' = I \end{aligned}$$

and

$$\begin{aligned} \boldsymbol{\theta}' &= \boldsymbol{\delta}'[I - B(B'B)^{-1}B']L \\ &= \boldsymbol{\delta}'[L_2L_2' + PP'][L_1 \vdots L_2] \\ &= \boldsymbol{\delta}'[PP'L_1 \vdots L_2 + PP'L_2] \\ &= \boldsymbol{\delta}'[0 \vdots L_2], \quad \text{since} \quad \boldsymbol{\delta}'P = 0 \end{aligned}$$

Thus as $n \to \infty$,

$$w_{i,n} \xrightarrow[p]{} 0, \text{ for } i = 1, 2, \cdots, s + q,$$

and $w_{s+q+1,n}, w_{s+q+2,n}, \cdots, W_{R,n}$ are asymptotically distributed as independent normal variates with unit variance and means given by

$$\lim_{n\to\infty} E(w_{s+q+1,n}, w_{s+q+2,n}, \cdots, w_{R,n}) = \boldsymbol{\delta}' L_2$$

Hence

$$\sum_{i=1}^{q} \sum_{j=1}^{r_i} \frac{(v_{ijn} - N_n Q_i p_{ij} (\hat{\alpha}_{1n}, \hat{\alpha}_{2n}, \cdots, \hat{\alpha}_{sn}))^2}{N_n Q_i p_{ij} (\hat{\alpha}_{1n}, \hat{\alpha}_{2n}, \cdots, \hat{\alpha}_{sn})}$$
$$= Y'_{(n)} Y_{(n)} = W'_{(n)} W_{(n)} = \sum_{i=1}^{R} w_{i,n}^2$$

is, in the limit as $n \to \infty$, distributed as non-central χ^2 with $R - s - q$ degrees of freedom and noncentrality parameter

$$\begin{aligned} \lambda &= \boldsymbol{\delta}' L_2 L_2' \boldsymbol{\delta} \\ &= \boldsymbol{\delta}'(PP' + L_2 L_2')\boldsymbol{\delta} \\ &= \boldsymbol{\delta}'(I - B(B'B)^{-1}B')\boldsymbol{\delta} \end{aligned}$$

This completes the proof of Theorem 3.1 (ii). It will be seen that the proof of Theorem 3.1 given here, follows reasoning similar to that in Cramér ([15], section 30.3). An alternative proof is also possible on the lines of Wald's derivation (Theorem IX [17]) of the large sample distribution of the likelihood ratio criterion, with suitable modifications.

4. The limiting power of the frequency χ^2-test. Neyman [13] considers the following problem:

Consider q sequences of independent trials and let $N_{(i)}$ denote the number of trials in the ith sequence. Each trial of the ith sequence is capable of producing one of the r_i mutually exclusive results, say

$$\rho_{i,1}, \quad \rho_{i,2}, \quad \cdots, \quad \rho_{i,r_i}$$

with unknown probabilities

$$\overset{\bullet}{p}_{i1}, \quad \overset{\bullet}{p}_{i2}, \quad \cdots, \quad \overset{\bullet}{p}_{ir_i}$$

where

$$\sum_{i=1}^{r_i} \overset{\bullet}{p}_{ij} = 1$$

Denote by v_{ij} the number of occurences of $\rho_{i,j}$ in the course of the $N_{(i)}$ trials forming the ith sequence.

On the basis of these observations $\{v_{ij}\}$ it is desired to test the hypothesis

that these unknown probabilities p^*_{ij} satisfy certain known functional relations, e.g.

$$H : p^*_{ij} = p_{ij}(\alpha_1, \alpha_2, \cdots, \alpha_s)$$

where the p_{ij}'s are certain functions satisfying the conditions described in section 3, and $(\alpha_1, \alpha_2, \cdots, \alpha_s)$ is an unknown parameter point. Let $\hat{\alpha}_1, \hat{\alpha}_2, \cdots, \hat{\alpha}_s$ be a suitably chosen solution of

$$\sum_{i=1}^{q} \sum_{j=1}^{r_i} \frac{v_{ij} - N_{(i)} p_{ij}}{p_{ij}} \frac{\partial p_{ij}}{\partial \alpha_k} = 0, \qquad k = 1, 2, \cdots, s. \tag{4.1}$$

and let $\chi^2_{1-\alpha}(u)$ be the upper α percent point of the χ^2-distribution with u degrees of freedom.

For testing H we compute

$$\chi^2_H = \sum_{i=1}^{q} \sum_{j=1}^{r_i} \frac{(v_{ij} - N_{(i)} p_{ij} (\hat{\alpha}_1, \hat{\alpha}_2, \cdots, \hat{\alpha}_s))^2}{N_{(i)} p_{ij} (\hat{\alpha}_1, \hat{\alpha}_2, \cdots, \hat{\alpha}_s)^2}$$

We reject H if $\chi^2_H > \chi^2_{1-\alpha}(R - s - q)$, and accept otherwise. Put $N = \sum_{i=1}^{q} N_{(i)}$ and $Q_i = N_{(i)}/N$. Let $\{c_{ij}\}$, $\mathbf{\delta}$ and B be as defined in section 3. Let $F(\chi^2, u, \lambda)$ be the distribution function of the non-central χ^2 with u degrees of freedom and non-centrality parameter λ. Define the hypothesis

$$H_N : p^*_{ij} = p_{ij} (\alpha^0_1, \alpha^0_2, \cdots, \alpha^0_s) + \frac{c_{ij}}{\sqrt{N}} = p_{ijN} \text{ (say)},$$

where as before $(\alpha^0_1, \alpha^0_2, \cdots, \alpha^0_s)$ is an inner point of A.

From Theorem 3.1, we obtain the limiting power of the χ^2_H-test

$$\beta(\chi^2_H, \{H_N\}) = 1 - F(\chi^2_{1-\alpha}(R - s - q), R - s - q, \lambda)$$

where $\lambda = \mathbf{\delta}'(I - B(B'B)^{-1}B)\mathbf{\delta}$.

Let $\mathbf{d}' = (d_1, d_2, \cdots, d_s)$ be any vector of real numbers. When

$$\{c_{ij}\}_{R \times 1} = D\mathbf{d},$$

it is easily seen that

$$p_{ij} (\alpha^0_1, \alpha^0_2, \ldots, \alpha^0_s) + \frac{c_{ij}}{\sqrt{N}} = p_{ij}(\alpha_{1N}, \alpha_{2N}, \ldots, \alpha_{sN}) + o\left(\frac{1}{\sqrt{N}}\right)$$

where $\alpha_{kN} = \alpha^0_k + d_k/N^{\frac{1}{2}}(k = 1, 2, \cdots, s)$. In this case $\mathbf{\delta}$ is of the form

$$\mathbf{\delta} = B \cdot \mathbf{e}$$

where $\mathbf{e}' = (e_1, e_2, \cdots, e_s)$ is another real vector. We have

$$\begin{aligned}\lambda &= \mathbf{e}'B'(I - B(B'B)^{-1}B')B\mathbf{e} \\ &= 0\end{aligned}$$

and $\beta(\chi^2_H, \{H_N\}) = \alpha$, as we might expect.

5. Applications. (1) *Planning of experiments for comparing two distribution functions.*

To test the hypothesis that two random variables x_1 and x_2 have identical probability distributions, the test procedure commonly adopted consists in making a sequence of N_i independent observations on the random variable x_i $(i = 1, 2)$. At each observation we observe the numerical value assumed by the random variable and according to this classify the results of each sequence into r measurable mutually exclusive and exhaustive groups (same for both the sequences).

Let v_{ij} denote the number of observations of the ith sequence belonging to the jth group $(i = 1, 2, j = 1, 2, \cdots, r)$, so that $\sum_{j=1}^{r} v_{ij} = N_i (i = 1, 2)$. The hypothesis desired to be tested is equivalent to the hypothesis H^* that there are r positive constants $p_1, p_2, \cdots, p_r$ with $\sum_{j=1}^{r} p_j = 1$ such that the probability of a random observation belonging to the jth group is equal to p_j for both the sequences. (We assume that the groups are so chosen that each of them has a positive probability measure at least w.r.t. one of the distributions.)

If this hypothesis H^* is true, the maximum likelihood estimates of p_j will be given by $\hat{p}_j = v_{.j}/N$, where $v_{.j} = v_{1j} + v_{2j}$ and $N = N_1 + N_2$. Hence for testing the hypothesis we compute

$$\chi^2_{H^*} = \sum_{i=1}^{2} \sum_{j=1}^{r} \frac{(v_{ij} - Q_i v_{.j})^2}{Q_i v_{.j}} \tag{5.1}$$

We reject the hypothesis if

$$\chi^2_{H^*} > \chi^2_{1-\alpha}(r - 1),$$

and accept it otherwise.

Let us now assume that it costs C_i dollars to make an observation on x_i $(i = 1, 2)$. Since both N_1 and N_2 are at our disposal it seems now natural to inquire how best we could allocate our total sampling budget of S dollars to the two populations, or, more precisely, could we determine the ratio $N_1 / (N_1 + N_2) = Q_1$ which will maximize the power of the above test with respect to all alternatives violating the hypothesis H^*, and at the same time ensure that the sampling cost does not exceed S dollars. Due to reasons already stated earlier in this paper, we cannot provide an answer to this question with our existing knowledge. However, if we agree to accept the limiting power function as our criterion for choosing 'the best', we might seek if the best possible sampling plan exists in the sense of maximizing the limiting power.

Let c_{ij} $(i = 1, 2, j = 1, 2, \cdots, r)$ be any given set of deviation parameters such that

$$\sum_{j=1}^{r} c_{ij} = 0, \qquad i = 1, 2, \text{ and for at least one } j,$$

$$c_{1j} \neq c_{2j}.$$

Let us denote by H_s^* the hypothesis

$$H_s^* : p_{ij}(S) = p_j^0 + \frac{c_{ij}}{\sqrt{S}}.$$

If we decide to take N_1 and N_2 in the ratio $Q_1 : (1 - Q_1)$ then the total sample size will be given by

$$N = \frac{S}{C_1 Q_1 + C_2 Q_2} \qquad \text{where} \quad Q_2 = 1 - Q_1 .$$

Hence H_s^* may be rewritten as

$$H_s^* : p_{ij}(S) = p_j^0 + \frac{c_{ij}}{\sqrt{C_1 Q_1 + C_2 Q_2}\sqrt{N}}.$$

From Theorem 3.1, we obtain the limiting power of the $\chi^2_{H^*}$-test

$$\beta(\chi^2_{H^*}, \{H_s^*\}) = 1 - F(\chi^2_{1-\alpha}(r-1), (r-1), \chi_{H^*})$$

where

$$\lambda_{H^*} = \bar{\delta}'(I - B(B'B)^{-1}B')\bar{\delta}.$$

After some simplification λ_{H^*} reduces to

$$\frac{Q_1 Q_2}{C_1 Q_1 + C_2 Q_2} \sum_{j=1}^{r} (c_{1j} - c_{2j})^2 / p_j^0$$

Since for given x and u, $F(x, u, \lambda)$ is a strictly monotonic decreasing function of λ, the maximum limiting power is attained when λ_{H^*} is maximum, that is when

$$Q_1 = \frac{\sqrt{C_2}}{\sqrt{C_1} + \sqrt{C_2}}$$

Thus to maximize the limiting power the best possible sampling plan, at the specified budget, is given by

$$N_1 = \left[\frac{S}{\sqrt{C_1}\,(\sqrt{C_1} + \sqrt{C_2})}\right]$$

and

$$N_2 = \left[\frac{S}{\sqrt{C_2}\,(\sqrt{C_1} + \sqrt{C_2})}\right]$$

where $[x]$ denotes the largest integer less than x.

(2) *Planning of experiments to detect shifts in response.*

Consider the following problem discussed by McNemar [18] who was interested in ascertaining the effectiveness of an interpolated experience like a movie or a lecture in shifting individual responses to certain stimuli. Let us take the simple

situation in which every individual responds to the stimuli in one of two different ways (say, '0' or '1'). Let π_{ij} denote the proportion of individuals in the population, who give response 'i' before the interpolated experience and response 'j' after it ($i = 0, 1; j = 0, 1$).

Write

$$\pi_{i.} = \pi_{i0} + \pi_{i1} \qquad (i = 0, 1)$$

$$\pi_{.j} = \pi_{0j} + \pi_{1j} \qquad (j = 0, 1)$$

We shall say there is no shift in response if

$$H_0 : \pi_{1.} = \pi_{.1}$$

is true.

To test this hypothesis one can conceive of at least two alternative ways of experimentation:

(a) two samples, each of size n are selected independently, one from the pre-experience group and the other from the post-experience group. The test for the equality of proportions then, is easily seen to be a particular case of the test given earlier in this section under Application (1). Let us denote the chisquare obtained for this test by χ_a^2.

(b) the same set of individuals, n, in number selected from the pre-experience group is again examined after the experience, and the results classified in a 2×2 table as follows:

	Post experience response		
	0	1	total
Pre-experience response			
0	n_{00}	n_{01}	$n_{0.}$
1	n_{10}	n_{11}	$n_{1.}$
total	$n_{.0}$	$n_{.1}$	n

Under procedure (b), to test H_0, we compute

$$\chi_b^2 = \frac{(n_{10} - n_{01})^2}{n_{01} + n_{10}}$$

and reject H_0, only if, $\chi_b^2 > \chi_{1-\alpha}^2$ (1). Let us denote by H_{0n} the hypothesis

$$H_{0n} : \pi_{ij} = \pi_{ij}^0 + \frac{c_{ij}}{\sqrt{n}},$$

where $\Sigma\pi_{ij}^0 = 1$, $\pi_{01}^0 = \pi_{10}^0 = \pi^0$ (say), $\Sigma c_{ij} = 0$, and $c_{01} \neq c_{10}$. From Theorem 3.1, we obtain after certain algebraic simplification:

$$\beta(\chi_a^2, \{H_{0n}\}) = 1 - F(\chi_{1-\alpha}^2(1), 1, \lambda_a)$$

and

$$\beta(\chi_b^2, \{H_{0n}\}) = 1 - F(\chi_{1-\alpha}^2(1), 1, \lambda_b),$$

where

$$\lambda_a = \frac{(c_{10} - c_{01})^2}{2\{(\pi_{11}^0 + \pi^0)(\pi_{00}^0 + \pi^0)\}}$$

and

$$\lambda_b = \frac{(c_{10} - c_{01})^2}{2\pi^0}.$$

The denominator in λ_a can be rewritten as $2\{\pi^0 - \pi^{0^2} + \pi_{00}^0\pi_{11}^0\}$. Hence, $\lambda_b >$, $<$ or $= \lambda_a$, according as $(\pi_{00}^0\pi_{11}^0 - \pi_{01}^0\pi_{10}^0) >, <$ or $= 0$ respectively. This shows that at least from the point of view of maximising limiting power, procedure (b) would be superior to procedure (a) when the association between the two response types, as measured by $\pi_{00}^0\pi_{11}^0 - \pi_{01}^0\pi_{10}^0$, is positive; inferior to (a) when it is negative; and equivalent to (a) when it is zero.

ACKNOWLEDGEMENT. I would like to thank Professor S. N. Roy for his continued interest in this work and his helpful suggestions, and the referees for their valuable comments.

REFERENCES

[1] C. EISENHART, "The power function of the χ^2 test," (Abstract) *Bull. Amer. Math. Soc.*, Ser. 2, Vol. 44 (1938), p. 32.

[2] P. B. PATNAIK, "The non-central χ^2 and F-distributions and their applications," *Biometrika*, Vol. 36 (1949), pp. 202–232.

[3] P. B. PATNAIK, "The power function of the test for the difference between two proportions in a 2×2 table," *Biometrika*, Vol. 35 (1948), pp. 157–175.

[4] G. P. SILLITTO, "Note on approximations to the power function of the '2×2 comparative trial'," *Biometrika*, Vol. 36 (1949), pp. 347–352.

[5] W. L. STEVENS, "Mean and variance of an entry in a contingency table," *Biometrika*, Vol. 38 (1951), pp. 468–470.

[6] E. S. PEARSON AND M. MERRINGTON, "2×2 Tables; the power function of the test on a randomized experiment," *Biometrika*, Vol. 35 (1948), pp. 331–345.

[7] S. J. POTI, "Power function of Chi-Square test with special reference to analysis of blood group data," *Sankhya*, Vol. 10 (1950), pp. 397–406.

[8] C. L. CHIANG, "On the Design of Mass Medical Surveys," *Human Biology*, Vol. 23 (1951), pp. 242–271.

[9] W. F. TAYLOR, "Mathematical Statistical Studies on the Tuberculin and Histoplasmin skin tests, I, overlapping of Negative and Positive Reactions," *Human Biology*, Vol. 23 (1951), pp. 1–23.

[10] W. G. COCHRAN, "The χ^2 test of goodness of fit," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 314–345.

[11] E. J. G. PITMAN, "Unpublished lecture notes on non-parametric statistical inference," 1948.

[12] W. HOEFFDING AND J. R. ROSENBLATT, "The Efficiency of Tests," *Ann. Math. Stat.*, Vol. 26 (1955), pp. 52–63.

[13] J. NEYMAN, "Contributions to the theory of χ^2 test," *Proceedings of the Berkeley Sym-*

posium on Mathematical Statistics and Probability, University of California Press, 1949.

[14] E. Fix, "Tables of Noncentral χ^2," *University of California Publications in Statistics*, Vol. 1, No. 2, pp. 15–19, University of California Press, 1949.

[15] H. Cramér, *Mathematical Methods of Statistics*, Princeton University Press, 1946.

[16] S. K. Mitra, *Contributions to the Statistical Analysis of Categorical Data*, Institute of Statistics Mimeograph Series, University of North Carolina, No. 142 (1955), pp. 1–146.

[17] A. Wald, "Tests of Statistical Hypotheses Concerning Several Parameters when the number of observations is large," *Trans. Amer. Math. Soc.*, Vol. 54, (1943), pp. 426–482.

[18] Q. McNemar, *Psychological Statistics*, 2nd edition, Chapman and Hall, London, 1955.

SOME EXACT RESULTS FOR THE FINITE DAM

By N. U. Prabhu

University of Western Australia, and Karnatak University, India

1. Summary. In the discrete finite dam model due to Moran, the storage process $\{Z_t\}$ is known to be a Markov chain. Stationary distributions of Z_t are obtained for the cases where the release is a unit amount of water per unit time, and the input is of (i) geometric, (ii) negative binomial and (iii) Poisson type.

The paper concludes with a discussion of the problem of emptiness in the finite dam and considers the probability that, starting with an arbitrary storage, the dam becomes empty before it overflows.

2. Introduction. This paper is concerned with a storage system whose probability model is due to Moran [9]. The storage Z_t of a dam of finite capacity K is defined for discrete time t ($t = 0, 1, 2, \cdots$) as the dam content just after an instantaneous release at t, and just before an input X_t flows into it over the time-interval $(t, t + 1)$. The model is subject to the conditions that

(i) the inputs X_t during the intervals $(t, t + 1)$ are independently and identically distributed;

(ii) there is an overflow $\text{Max}(Z_t + X_t - K, 0)$ during the interval $(t, t + 1)$, a quantity $\text{Min}(K, Z_t + X_t)$ being left in the dam just before the release occurs; and

(iii) the amount of water released at time $t + 1$ is $\text{Min}(M, Z_t + X_t)$ where M is a constant ($<K$).

A fuller description of the model and further references on the subject are given by Gani [3]. It is seen the stochastic processes $\{Z_t\}$ and $\{Z_t + X_t\}$ are both Markov chains, and the problem of obtaining their stationary distributions, given the probability distribution of the input, is of some interest. Moran ([9], [10]) and Gani and Moran [4] have obtained a few approximate solutions to this problem by numerical methods, and some important observations on the solution in the general case have been made by Moran [11], but the only exact solution known so far is the one due to Moran [10] for the case of the geometric input. The problem is considerably simplified when $K = \infty$, i.e. when the dam is of infinite capacity; it is then seen (Gani and Prabhu, [5]) that the transition-matrix of the Markov chain $\{Z_t + X_t\}$ also occurs in the theory of queues in connection with the length of a queue at epochs just before service. For this case Bailey [1] has obtained, by the method of probability generating functions (p.g.f.), the stationary distributions arising from a given distribution of X_t. A dam of finite capacity K can be considered as the analogue of a queueing system in which there is accommodation for only K customers to wait, those in excess of K being compelled to leave the queue altogether (as may happen, for

Received November 8, 1957; revised March 27, 1958.

instance, in an airport of limited capacity); we proceed to obtain for such a dam the stationary distribution of the storage Z_t.

3. Stationary distribution of the storage. We shall be concerned with the case where M, the amount of water released at time t, is unity. Let $\{g_j\}$ be the probability distribution of X_t, so that

$$\Pr\{X_t = j\} = g_j, \qquad (j = 0, 1, 2, \cdots). \tag{1}$$

We assume that $g_j > 0$ for all j. Also, let

$$G(z) = \sum_{j=0}^{\infty} g_j z^j, \qquad |z| < 1 \tag{2}$$

be the p.g.f. of $\{g_j\}$, and

$$\rho = G'(1) = \sum_{j=0}^{\infty} j g_j \tag{3}$$

the mean input. The transition-matrix of the Markov chain $\{Z_t\}$ is $P \equiv \{P_{ij}\}$, where

$$P = \begin{array}{c|cccccc} \nearrow & 0 & 1 & \cdots & K-2 & K-1 \\ \hline 0 & g_0+g_1 & g_2 & \cdots & g_{K-1} & h_K \\ 1 & g_0 & g_1 & \cdots & g_{K-2} & h_{K-1} \\ 2 & 0 & g_0 & \cdots & g_{K-3} & h_{K-2} \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\ K-1 & 0 & 0 & \cdots & g_0 & h_1 \end{array} \tag{4}$$

where $h_i = \sum_{j=i}^{\infty} g_j$, $(i = 1, 2, \cdots, K)$. Clearly, the chain is irreducible and contains a finite number K of states, so that the stationary probability distribution $\{u_i\}$, $(i = 0, 1, \cdots, K-1)$ exists, where the u_i are the unique solutions of the equations

$$u_j = \sum_{i=0}^{K-1} u_i p_{ij}, \qquad (j = 0, 1, \cdots, K-1) \tag{5}$$

together with $\sum_0^{K-1} u_i = 1$. We first prove the following theorem due to Moran [11].

THEOREM.

(i) *If* $\{u_i^{(K)}\}$, $(i = 0, 1, \cdots, K-1)$ *is the stationary probability distribution of storage in a dam of capacity* K, *then the ratios*

$$v_i = \frac{u_i^{(K)}}{u_0^{(K)}}, \qquad (i = 1, 2, \cdots, K-1) \tag{6}$$

are independent of K, *and*

(ii) *the v_i's can be found as the coefficients of z^i in $V(z)$, where*

$$v(z) = \frac{g_0(1 - z)}{G(z) - z} \tag{7}$$

The first part of the theorem is easily proved; in fact, writing out the equations (5) in full we obtain

$$\begin{aligned} u_0 &= (g_0 + g_1)u_0 + g_0 u_1 \\ u_1 &= g_2u_0 \quad + g_1u_1 + g_0u_2 \\ &\cdot \quad \cdot\ \cdot \quad\quad \cdot\ \cdot \quad\quad \cdot\ \cdot \\ u_{K-2} &= g_{K-1}u_0 + g_{K-2}u_1 + \cdots + g_0u_{K-1} \\ u_{K-1} &= h_Ku_0 \quad + h_{K-1}u_1 + \cdots + h_1u_{K-1} \end{aligned}$$

Solving these equations successively for the ratios $v_i = u_i / u_0$ we obtain

$$\begin{aligned} v_1 &= \frac{1 - g_0 - g_1}{g_0} \\ v_2 &= \frac{1 - g_1}{g_0} v_1 + \frac{g_2}{g_0} \end{aligned} \tag{8}$$

and in general, the v_i's ($i = 1, 2, \cdots, K - 1$) are seen to be independent of K. Now consider the function $V(z)$ defined by (7). We shall first prove that $V(z)$ can be expanded as a power series which is convergent for suitable values of $|z|$. Let us first consider the case $\rho \leqq 1$. Writing

$$G(z) - z = (1 - z)\left\{1 - \frac{1 - G(z)}{1 - z}\right\}$$

and following Kendall ([6], p. 159) we obtain

$$\frac{1 - G(z)}{1 - z} = \sum_{n=0}^{\infty} z^n \sum_{n+1}^{\infty} g_i$$

so that, for $|z| < 1$,

$$\left|\frac{1 - G(z)}{1 - z}\right| < \sum_{n=0}^{\infty} \sum_{n+1}^{\infty} g_i = \sum_{i=1}^{\infty} ig_i = \rho \leqq 1$$

Hence $|G(z) - z| \neq 0$, and we have the power series expansion

$$V(z) = g_0\left\{1 - \frac{1 - G(z)}{1 - z}\right\}^{-1} = v_0 + v_1 z + v_2 z^2 + \cdots$$

convergent for $|z| < 1$.

Next, let $\rho > 1$. In this case there exists a positive λ such that the power series expansion

$$\frac{1}{G(z) - z} = c_0 + c_1 z + c_2 z^2 + \cdots$$

is valid for $|z| < \lambda$ (Knopp, [8], p. 182). Hence it follows that $V(z)$ also possesses a power series expansion convergent for $|z| < \lambda$.

Thus whether or not $\rho \leqq 1$, $V(z)$ has a power series expansion

$$V(z) = \frac{g_0(1-z)}{G(z) - z} = v_0 + v_1 z + v_2 z^2 + \cdots \tag{9}$$

The coefficients v_i are determined from the relation

$$g_0(1 - z) = \{G(z) - z\} \sum_{i=0}^{\infty} v_i z^i$$

and hence it is seen that $v_0 = 1$, and $v_1, v_2, \cdots, v_{K-1}$ satisfy the relations (8). Thus they are, in fact, the quantities defined in (6).

If $\rho < 1$, the stationary probability distribution exists in the case of the infinite chain ($K = \infty$), and its g.f. is proportional to $V(z)$. However, the above results hold, as we have shown, even when $\rho \geqq 1$. It is now obvious that the general method of obtaining the stationary probability distribution $\{u_i\}$ for the discrete dam of finite capacity K consists of (i) finding $V(z)$, (ii) expanding $V(z)$ to obtain the v_i's, and (iii) normalising $v_0, v_1, \cdots, v_{K-1}$ to obtain a probability distribution. We proceed to do this in some particular cases.

3.1. Geometric input. Consider, for instance, an input distribution of the geometric type,

$$g_j = \Pr\{X_t = j\} = ab^j, \qquad (j = 0, 1, \cdots,) \tag{10}$$

where $0 < a < 1$ and $b = 1 - a$. The p.g.f. of X_t is then

$$G(z) = \frac{a}{1 - bz} \tag{11}$$

and the function $V(z)$ is given by

$$\begin{aligned} V(z) &= \frac{a(1-z)}{a(1-bz)^{-1} - z} = \frac{1 - bz}{1 - \rho z} \\ &= (1 - bz) \sum_{i=0}^{\infty} \rho^i z^i \left(|z| < \min\left(\frac{1}{\rho}, 1\right)\right), \end{aligned} \tag{12}$$

where $\rho = b/a$ is the mean input. Hence we obtain

$$v_0 = 1, \qquad v_i = \rho^i - b\rho^{i-1} = b\rho^i, \qquad (i = 1, 2, \cdots, K-1)$$

and

$$\sum_0^{K-1} v_i = 1 + b \sum_{i=1}^{K-1} \rho^i = a \frac{1 - \rho^{K+1}}{1 - \rho}.$$

The stationary distribution in this case is therefore given by $\{u_i\}$, where

$$u_0 = \frac{(1-\rho)}{a(1 - \rho^{K+1})}, \qquad u_i = \frac{\rho^{i+1}(1-\rho)}{1 - \rho^{K+1}}, \qquad (i = 1, 2, \cdots, K-1). \tag{13}$$

Thus the storage of a dam of finite capacity K into which flows an input of the geometric type has a stationary distribution of the geometric type, which is truncated at $Z = K - 1$ and has a modified initial term. This result is implied in Moran's solution (referred to in Section 2) for the general case $M > 1$, although it is not explicitly mentioned by him; for $M = 1$ his solution is given by the formulae $u_0 = \pi_0 + \pi_1$, $u_i = \pi_{i+1}$ $(i = 1, 2, \cdots, K - 1)$, where

$$\pi_{K-r}/\pi_K = {}^rS_1\, a - {}^{r-1}S_2\, a^2 + {}^{r-2}S_3\, a^3 \cdots$$

$$(14) \qquad {}^nS_p = \binom{n-1}{p-1} b^{-n} - \binom{n-2}{p-1} b^{1-n}. \qquad (r = 1, 2, \cdots, K)$$

From this we obtain (13) after some simple reduction.

3.2. Negative binomial input. Consider next the more general case of the negative binomial input,

$$(15) \qquad g_j = \Pr\{X_t = j\} = n_j \binom{n+j-1}{j} a^n b^j, \qquad (j = 0, 1, \cdots)$$

where $0 < a < 1$, $b = 1 - a$, and n is a positive integer; the p.g.f. of X_t is then

$$G(z) = \frac{a^n}{(1 - bz)^n}$$

and the mean input is $\rho = nb/a$. We have then

$$(16) \qquad V(z) = \frac{a^n(1 - z)}{a^n(1 - bz)^{-n} - z} = \frac{a^n(1 - z)(1 - bz)^n}{a^n - z(1 - bz)^n}$$

Obviously $z = 1$ is a zero of the denominator of the expression on the right hand side of (16); in addition to this it has n other zeros $z_1, z_2, \cdots, z_n$. We consider here the case where $z_1, z_2, \cdots, z_n$ are all distinct and different from unity; however, the general case can be treated along similar lines. When $(1, z_1, z_2, \cdots, z_n)$ are all different we can break up $V(z)$ into partial fractions of the form

$$(17) \qquad V(z) = d_0 + \sum_{p=1}^{n} \frac{d_p}{1 - z/z_p}$$

where obviously $d_0 = a^n$ and the d_p's are given by

$$d_p = \lim_{z \to z_p} \left(1 - \frac{z}{z_p}\right) V(z)$$

$$(18) \qquad = \lim_{z \to z_p} \frac{a^n(1 - z)(1 - bz)^n \left(1 - \dfrac{z}{z_p}\right)}{a^n - z(1 - bz)^n} = \frac{a^n(1 - 1/z^p)}{\rho a z_p/(1 - bz_p) - 1}$$

$$(p = 1, 2, \cdots, n).$$

Now let λ be the least among the quantities $1, |z_1|, |z_2|, \cdots, |z_n|$; then for $|z| < \lambda$ we can express each term under the summation sign in (17) as a power series. Thus

$$V(z) = d_0 + \sum_{p=1}^{n} d_p \sum_{i=0}^{\infty} \left(\frac{z}{z_p}\right)^i = d_0 + \sum_{i=0}^{\infty} z^i \sum_{p=1}^{n} d_p \left(\frac{1}{z_p}\right)^i,$$

whence we obtain

$$(19) \qquad \begin{aligned} v_0 &= d_0 + \sum_{p=1}^{n} d_p = \lim_{z \to 0} V(z) = 1 \\ v_i &= \sum_{p=1}^{n} \frac{d_p}{(z_p)^i}, \qquad (i = 1, 2, \cdots, K-1), \end{aligned}$$

so that

$$\sum_{i=0}^{K-1} v_i = d_0 + \sum_{p=1}^{n} d_p \sum_{i=0}^{K-1} \left(\frac{1}{z_p}\right)^i = d_0 + \sum_{p=1}^{n} d_p \frac{1 - (1/z_p)^K}{1 - (1/z_p)}.$$

It follows that the stationary probabilities u_i are given by

$$(20) \qquad \begin{aligned} u_0 &= \left\{ d_0 + \sum_{p=1}^{n} d_p \frac{1 - (1/z_p)^K}{1 - (1/z_p)} \right\}^{-1} \\ u_i &= u_0 \sum_{p=1}^{n} d_p \left(\frac{1}{z_p}\right)^i, \qquad (i = 1, 2, \cdots K-1). \end{aligned}$$

From (20) we see that the stationary distribution of the dam storage is the weighted sum of n geometric distributions, each of which is truncated at $Z = K - 1$, and has a modified initial term.

3.3. Poisson input. Finally we consider the case where the input has the Poisson distribution with mean ρ,

$$(21) \qquad g_j = \Pr\{X_t = j\} = e^{-\rho} \frac{\rho^j}{j!}, \qquad (j = 0, 1, \cdots).$$

The rigorous procedure here consists of writing down $V(z)$ and obtaining the coefficients v_i by complex variable methods. We shall, however, argue heuristically and consider (21) as the limiting case of the negative binomial (15) as $n \to \infty$, $a \to 1$ and $\rho = nb / a$ is held fixed. In fact, putting $a = 1 / (1 + \rho/n)$, $b = \rho/(n + \rho)$, it is seen that the p.g.f. of (15) reduces to

$$(1 + \rho/n)^{-n} \left\{ 1 - \frac{1}{n} \rho \left(1 + \frac{1}{n} \rho\right)^{-1} z \right\}^{-n} \to e^{-\rho + \rho z}$$

which is the p.g.f. of (21). Also, $d_0 \to e^{-\rho}$, and

$$d_p = \frac{a_n(1 - 1/z_p)}{\rho a z_p/(1 - bz_p) - 1} \to \frac{e^{-\rho}(1 - 1/z_p)}{\rho z_p - 1}$$

where $z_1, z_2, \cdots$ are the roots (other than unity) of the equation

$$e^{-\rho+\rho z} = z \tag{22}$$

(which are infinite in number). Hence the stationary probabilities of the dam storage are given by

$$\begin{aligned} u_0 &= \left\{e^{-\rho} + \sum_{p=1}^{\infty} \frac{e^{-\rho}(1-1/z_p)}{\rho z_p - 1} \frac{1-(1/z_p)^K}{1-(1/z_p)}\right\}^{-1} \\ u_i &= u_0 \sum_{p=1}^{\infty} \frac{e^{-\rho}(1-1/z_p)}{\rho z_p - 1} \left(\frac{1}{z_p}\right)^i, \qquad (i = 1, 2, \cdots, K-1). \end{aligned} \tag{23}$$

4. The problem of emptiness in the finite dam. The analogy between the dam process and the random walk has already been pointed out by several authors (see the discussion in [3]). In fact, putting $U_t = X_t - 1$, we see that the storage Z_t in a dam of capacity K satisfies the relations

$$Z_{t+1} = \begin{cases} Z_t + U_t & \text{if } 0 < Z_t + U_t < K-1 \\ 0 & \text{if } \quad Z_t + U_t \leqq 0 \\ K-1 & \text{if } \quad Z_t + U_t \geqq K-1 \end{cases} \tag{24}$$

which, however, define a random walk with impenetrable barriers at $Z = 0$ and $Z = K - 1$. If $K = \infty$, there is only the first barrier and the problem of 'duration of the game' (i.e. the distribution of time required for the dam to become empty for the first time) has been discussed by Kendall [7] for the case where the input is of the Gamma type and the release is continuous. For finite K this problem is much more difficult; however, for this case we propose to discuss the probability of absorption at $Z = 0$ (i.e. the conditional probability V_i that, starting with a storage $Z_0 = i$, the dam becomes empty ($Z_t = 0$) before it overflows). This is a familiar problem in random walk theory, and has been discussed, for instance, by Feller ([2], pp. 300–303); it is seen that the probabilities V_i $(i = 1, 2, \cdots, K-2)$ satisfy the relations

$$\begin{aligned} V_1 &= \sum_{j=1}^{K-2} g_j V_j + g_0 \\ V_i &= \sum_{j=i-1}^{K-2} g_{j-i+1} V_j, \qquad (i = 2, 3, \cdots, K-2). \end{aligned}$$

These equations simplify to some extent if we note that the states 0 and $K - 1$ are absorbing, so that $V_0 = 1$, $V_{K-1} = 0$; for we can then write

$$V_i = \sum_{j=i-1}^{K-1} g_{j-i+1} V_j, \qquad (i = 1, 2, \cdots, K-2). \tag{25}$$

Clearly, the coefficients on the right hand side of these equations correspond to the rows of the transition-matrix (4). It will now be found easiest to start at the

bottom right hand corner and work up to the left: thus

$$g_0V_{K-3} + g_1V_{K-2} + h_2\cdot 0 = V_{K-2}$$

so that

$$V_{K-3} = \frac{1 - g_1}{g_0} V_{K-2},$$

and similarly

$$V_{K-4} = \frac{(1 - g_1)V_{K-3} - g_2 V_{K-2}}{g_0},$$

etc. This shows that the ratios of the quantities

$$w_i = V_{K-1-i} \tag{26}$$

are again independent of K ($w_0 = 0$, $w_{K-1} = 1$); rewriting the equations (25) in terms of these quantities we obtain

$$w_i = \sum_{j=0}^{i} g_j w_{i-j+1}, \qquad (i = 1, 2, \cdots, K - 2). \tag{27}$$

Consider the system of equations (27) for $i = 1, 2, \cdots$ ad. inf., and put

$$W(z) = \sum_{i=1}^{\infty} \frac{w_i}{w_1} z^{i-1} \tag{28}$$

we have

$$\begin{aligned} zW(z) &= \sum_{i=1}^{\infty} \frac{z^i}{w_1} \sum_{j=0}^{i} g_j w_{i-j+1} \\ &= \sum_{j=1}^{\infty} g_j \sum_{i=j}^{\infty} \frac{w_{i-j+1}}{w_1} z^i + g_0 \sum_{i=1}^{\infty} \frac{w_{i+1}}{w_1} z^i \\ &= \sum_{j=1}^{\infty} g_j \sum_{i=1}^{\infty} \frac{w_i}{w_1} \cdot z^{1+j-1} + g_0 \sum_{i=2}^{\infty} \frac{w_i}{w_1} z^{i-1} \\ &= G(z)W(z) - g_0, \end{aligned}$$

whence we obtain the relation

$$W(z) = \frac{g_0}{G(z) - z}. \tag{29}$$

Folowing the same lines of argument as for $V(z)$, we can prove that $W(z)$ can be expanded as a power series convergent for suitable values of $|z|$. Let $W(z) = \sum_{i=0}^{\infty} \omega_{i+1} z^i$; then since $\omega_{K-1} = w_{K-1} / w_1 = 1/w_1$, we must have

$$w_i = \frac{\omega_i}{\omega_{K-1}}, \qquad (i = 1, 2, \cdots, K - 2) \tag{30}$$

which are, therefore, the required solutions to the equations (27). The absorption probabilities V_i can then be obtained from (26).

Let us now consider the particular case where the input is geometric with probabilities $g_j = ab^j$, $(j = 0, 1, 2, \cdots)$, and $G(z) = a(1 - bz)^{-1}$; then (29) gives

$$W(z) = \frac{a}{a(1-bz)^{-1} - z} = \frac{(1-bz)}{(1-z)(1-\rho z)}$$

$$(31) \qquad = \begin{cases} \dfrac{a}{1-\rho}\left\{\dfrac{1}{1-z} - \dfrac{\rho^2}{1-\rho z}\right\} & \text{if } \rho \neq 1 \\[2ex] \dfrac{1-bz}{(1-z)^2} & \text{if } \rho = 1. \end{cases}$$

Hence it follows that

$$(32) \qquad \omega_i = \begin{cases} \dfrac{a(1-\rho^{i+1})}{1-\rho} & \text{if } \rho \neq 1 \\[2ex] a(i+1) & \text{if } \rho = 1 \end{cases} \qquad (i \neq 1)$$

and

$$(33) \qquad w_i = \begin{cases} \dfrac{1-\rho^{i+1}}{1-\rho^K} & \text{if } \rho \neq 1 \\[2ex] \dfrac{i+1}{K} & \text{if } \rho = 1 \end{cases} \qquad (i = 1, 2, \cdots, K-2).$$

Thus the absorption probabilities V_i in the case of the geometric input are given by

$$(34) \qquad V_i = \begin{cases} \dfrac{1-\rho^{K-i}}{1-\rho^K} & \text{if } \rho \neq 1 \\[2ex] 1 - \dfrac{i}{K} & \text{if } \rho = 1 \end{cases} \qquad (i = 1, 2, \cdots, K-2).$$

A similar procedure could be used, when the input is of a more general type, to obtain the exact expressions for the probabilities V_i. However, in many cases, it may suffice to know the bounds within which V_i lie, and these bounds are given by Feller ([2], inequalities 8.11 and 8.12 on p. 303). In fact, noting that $E(U_t) = E(X_t - 1) = \rho - 1$, where ρ is the mean input, we have that

$$(35) \qquad \begin{aligned} &\frac{z_0^{K-1} - z_0^i}{z_0^{K-1} - 1} \leq V_i \leq 1 && \text{if } \rho < 1 \\ &\frac{z_0^i - z_0^{K-1}}{1 - z_0^{K-1}} \leq V_i \leq z_0^i && \text{if } \rho > 1 \\ &1 - \frac{i}{K-1} \leq V_i \leq 1 && \text{if } \rho = 1 \end{aligned}$$

where z_0 is the unique positive root (other than unity) of the equation $\sum_j z^j \Pr\{U_t = j\} = 1$, i.e. $\sum_{j=0}^{\infty} g_j z^j = z$, and $z_0 \gtrless 1$ according as $\rho \lessgtr 1$.

5. Concluding remarks and acknowledgements. When the input X_t has a continuous probability distribution, it is seen that the stationary distribution function of $Z_t + X_t$ satisfies an integral equation, which has been solved by the author in a recent paper (Prabhu, [12]) for the special case when the input distribution is of the Gamma type. A more realistic problem on which some work is in progress at the moment is the one dealing with the finite dam process in continuous time; however, our solutions for discrete time may be taken as useful approximations to this continuous case.

I am indebted to Dr. J. Gani and the referee for many helpful suggestions.

REFERENCES

[1] N. T. J. Bailey, "On queueing processes with bulk service," *J. Roy. Stat. Soc.*, B, Vol. 16 (1954), pp. 80–87.

[2] W. Feller, "An Introduction to Probability Theory and its Applications," John Wiley and Sons, 1950.

[3] J. Gani, "Problems in the probability theory of storage systems," *J. Roy. Stat. Soc.*, B, Vol. 14 (1957), pp. 181–207.

[4] J. Gani and P. A. P. Moran, "A solution of dam equations by Monte Carlo methods," *Aust. J. Appl. Sci.*, Vol. 6 (1955), pp. 267–273.

[5] J. Gani and N. U. Prabhu, "Stationary distributions of the negative exponential type for the infinite dam," *J. Roy. Stat. Soc.*, B Vol. 19 (1957), pp. 342–351.

[6] D. G. Kendall, "Some problems in the theory of queues," *J. Roy. Stat. Soc.*, B. Vol. 13 (1951), pp. 151–173.

[7] D. G. Kendall, "Some problems in the theory of dams," *J. Roy. Stat. Soc.*, B Vol. 19 (1957), pp. 207–212.

[8] K. Knopp, "Theory and applications of infinite series," Blackie and Son, London and Glasgow, 1928.

[9] P. A. P. Moran, "A probability theory of dams and storage systems," *Aust. J. Appl. Sci.*, Vol. 5 (1954), pp. 116–124.

[10] P. A. P. Moran, "A probability theory of dams and storage systems: modifications of the release rules," *Aust. J. Appl. Sci.*, Vol. 6 (1955), pp. 117–130.

[11] P. A. P. Moran, "A probability theory of dams with a continuous release," *Quart. J. of Math.*, (Oxford, 2), Vol. 7, pp. 130–137.

[12] N. U. Prabhu, "On the integral equation for the finite dam," *Quart. J. of Math.*, (Oxford, 2), (to appear).

MINIMAX ESTIMATION FOR LINEAR REGRESSIONS[1]

By R. Radner

University of California, Berkeley

1. Introduction and Summary. When estimating the coefficients in a linear regression it is usually assumed that the covariances of the observations on the dependent variable are known up to multiplication by some common positive number, say c, which is unknown. If this number c is known to be less than some number k, and if the set of possible distributions of the dependent variable includes "enough" normal distributions (in a sense to be specified later) then the minimum variance linear unbiased estimators of the regression coefficients (see [1]) are minimax among the set of all estimators; furthermore these minimax estimators are independent of the value of k. (The risk for any estimator is here taken to be the expected square of the error.) This fact is closely related to a theorem of Hodges and Lehmann ([3], Theorem 6.5), stating that if the observations on the dependent variable are assumed to be independent, with variances not greater than k, then the minimum variance linear estimators corresponding to the assumption of equal variances are minimax.

For example, if a number of observations are assumed to be independent, with common (unknown) mean, and common (unknown) variance that is less than k; and if, for every possible value of the mean, the set of possible distributions of the observations includes the normal distribution with that mean and with variance equal to k; then the sample mean is the minimax estimator of the mean of the distribution.

The assumption of independence with common unknown variance is, of course, essentially no less general than the assumption that the covariances are known up to multiplication by some common positive number, since the latter situation can be reduced to the former by a suitable rotation of the coordinate axes (provided that the original matrix of covariances is non-singular).

This note consideres the problem of minimax estimation, in the general "linear regression" framework, when less is known about the covariances of the observations on the "dependent variable" than in the traditional situation just described. For example, one might not be sure that these observations are independent, nor feel justified in assuming any other specific covariance structure. It is immediately clear that, from a minimax point of view, one cannot get along without any prior information at all about the covariances, for in that case the risk of every estimator is unbounded. In practice, however, one is typically willing to grant that the covarainces are bounded somehow, but one may not

Received June 4, 1957; revised May 5, 1958.

[1] Research undertaken by the Cowles Commission for Research in Economics under Contract Nonr-358(01), NR 047-006 with the Office of Naval Research.

have a very precise idea of the nature of the bound. One is therefore led to look for different ways of bounding the covariances, in the hope that the minimax estimators are not too sensitive to the bound.

Unfortunately, in the directions explored here, the minimax estimator is sensitive to the "form" of the bound, although once the form has been chosen the minimax estimator does not depend on the "magnitude" of the bound. This result thus provides an instance in which the minimax principle is not too effective against the difficulties due to vagueness of the statistical assumptions of a problem, although this is a type of situation in which it has often been successful (see Savage in [4], pp. 168–9).

In this note, two ways of bounding the covariances are considered. The first is equivalent to choosing a coordinate system for the "dependent variables," and placing a bound on the characteristic roots of the matrix of covariances of the coordinates, in terms of one of a certain class of metrics (e.g., placing a bound on the trace on the covariance matrix, or on its largest characteristic root). The second way consists of choosing a coordinate system, and then placing a bound on the variance of each coordinate.

In the first situation, the minimum variance linear unbiased estimator corresponding to the case of uncorrelated coordinates, with equal variances, turns out to be minimax; this minimax estimator is, in general, different for different choices of coordinate system, but does not depend on the "magnitude" of the bound. Also, the minimax loss typically decreases at the rate of the reciprocal of the sample size.

In the second situation, the minimax procedures derived here involve ignoring most of the observations, and applying a linear unbiased estimator to the rest. Again, the minimax procedure depends upon the choice of coordinate system; furthermore, in this case the minimax loss typically either does not approach zero with increasing sample size, or does so much more slowly than the reciprocal of the sample size.

Thus the minimax estimator appears to be less unsatisfactory in the first situation than in the second, but in both cases it depends upon the choice of coordinate system, which is a disadvantage if there is no "natural" coordinate system intrinsic to the regression problem being considered.

Section 2 below presents the formulation of the problem, and a basic lemma. Sections 3 and 4 explore the two ways of bounding the covariances just mentioned. Some examples are given in Section 5. I am indebted to R. R. Bahadur, L. J. Savage, and G. Debreu for their helpful comments.

2. Problem formulation and a basic lemma. Let y be a random N-dimensional column vector, with a distribution p that is known to be in some family P of distributions. Let $m_p = Ey$ denote the mean of the distribution p, and suppose that one is required to estimate the value of $f'm_p$ on the basis of a single observation on y, where f is given. It is assumed that the loss due to incorrect estimation is the square of the error. In this note minimax estimators of $f'm_p$ will be de-

rived under two different assumptions about P; both assumptions have the following form:

Let T be given $N \times M$ matrix; let C_p denote the covariance of p, i.e.,

$$C_p = E(y - m_p)(y - m_p)';$$

and let H be a given set of $N \times N$ covariance matrices.

(2.1) For every p in P, the mean $m_p = Tx$ for some M-dimensional vector x, and C_p is in H.

(2.2) For every x, and every C in H, there is a *normal* distribution in P with mean Tx and covariance C.

The assumption that P includes normal distributions is a natural one, since normality can rarely be ruled out as preposterous.

If α is any estimator, then the risk, or expected loss, associated with using α is, for any p, given by

$$(2.3) \qquad \begin{aligned} r(\alpha, p) &= E[\alpha(y) - f'm_p]^2 \\ &= E[\alpha(y) - E\alpha(y)]^2 + [E\alpha(y) - f'm_p]^2. \end{aligned}$$

An estimator $\hat{\alpha}$ is minimax if, for every estimator α,

$$\sup_{p \epsilon P} r(\hat{\alpha}, p) \leqq \sup_{p \epsilon P} r(\alpha, p).$$

Because of the convexity of the risk function, it is not necessary to consider randomized estimators (see [3], Theorem 3.2).

Relative to a given covariance C, an estimator α is said to be *minimum variance linear unbiased*, or more briefly, *Markoff*, if

$$(2.4) \qquad \alpha(y) = a'y \quad \text{(linearity)}.$$

$$(2.5) \qquad \text{For every } p \text{ in } P,\ Ea'y = m_p \quad \text{(unbiasedness)}.$$

(2.6) If β is any estimator satisfying (2.4) and (2.5), then for every p in P with covariance C, $r(\alpha, p) \leqq r(\beta, p)$

The significance of the Markoff estimators in this problem is that, in both cases considered in this note, there is a Markoff estimator, relative to some C in H, that is minimax.

It follows from (2.1) that a linear estimator a is unbiased if and only if $T'a = T'f$; and from (2.3) that the risk for a linear unbiased estimator is $a'C_pa$. Therefore, a linear estimator a is Markoff relative to C if and only if it minimizes $a'Ca$ subject to the constraint $T'a = T'f$.

It might be noted here that it follows from (2.3) that the standard definition of a Markoff estimator given above is equivalent to another one in which condition (2.5) (unbiasedness) is replaced by the following (bounded risk):

(2.5′) The risk $E(a'y - f'm_p)^2$ is bounded as p varies in the class of all p in P that have covariance C, for any given C.

The idea of replacing the constraint of unbiasedness by the constraint of bounded risk is close to the minimax spirit, and seems to be due to L. J. Savage.

The main tool that will be used is the following lemma, which is closely related to a theorem of Hodges and Lehman ([3], theorem 6.5), and is stated here without proof.

LEMMA. *If $\hat{a}$ is Markoff relative to $\hat{C}$ in H, and if $\hat{a}'C\hat{a} \leqq \hat{a}'\hat{C}\hat{a}$ for every C in H, then $\hat{a}$ is minimax.*

In the "classical" situation to which the general Markoff theorem on least squares is applied (see, for example, Aitken [1]), it is assumed that the covariance of the distribution p is known up to multiplication by a positive constant, i.e., that the covariance is cC, where C is known but c is not. If it is further assumed that c is bounded by some number k, then it follows immediately from the Lemma that the Markoff estimator relative to kC is minimax. Note that the Markoff estimator is independent of k.

On the other hand, if nothing at all is known about the covariance of p, i.e., if H is taken to be the class of *all* $N \times N$ covariance matrices, then the risk for every estimator is unbounded. To get a finite minimax value, the class H must be "bounded" in some sense, and the next two sections explore two directions in which such a bound can be defined. In each case it should be borne in mind that postulated assumptions are thought of as applying *after, possibly, an appropriate transformation of the coordinate system.*

3. The case of bounds in terms of characteristic roots. In this section minimax estimators are derived for the problem formulated in Section 2, when the covariances are bounded in certain ways in terms of their characteristic roots.

For any covariance matrix C, let r_i denote its characteristic roots (these will be non-negative real numbers). For any number $q \geqq 1$, the *q-norm* of C is defined here to be

$$N(C; q) = \left(\sum_i r_i^q\right)^{1/q}.$$

For $q = 1, 2$, and ∞, one gets the trace of C, the square root of the sum of squares of the elements of C, and the largest characteristic root of C, respectively. Note that for the identity matrix I, $N(I; q) = N^{1/q}$.

THEOREM 1. *Let q and k be given such that $1 \leqq q \leqq \infty$ and $k > 0$, and let H be the set of all covariances C such that $N(C; q) \leqq k$; then for the estimation problem described in Section 2, the Markoff estimator $\hat{a}$ relative to the identity matrix*[2] *is minimax, and the minimax loss is $k\hat{a}'\hat{a}$.*

PROOF. The idea of the proof is to show that the covariance of rank one that

[2] Strictly speaking, relative to the identity times an appropriate constant, since the identity may not be in H.

concentrates all the variance in the direction of $f'y$ is least favorable. Let $B = \hat{a}\hat{a}'/\hat{a}'\hat{a}$. Note that $N(B; q) = 1$. Since $\hat{a}$ is that unbiased linear estimator with minimum length, any unbiased linear estimator is of the form $\hat{a} + d$, where $\hat{a}'d = 0$. Hence for all unbiased linear estimators b,

$$b'Bb = \hat{a}'B\hat{a} = \hat{a}'\hat{a};$$

in particular, $\hat{a}$ is Markoff with respect to B, and to kB.

Let C be any covariance in H, and let r be its largest characteristic root; then

$$\hat{a}'C\hat{a} \leqq r\hat{a}'\hat{a} = N(C; \infty)\hat{a}'\hat{a} \leqq N(C; q)\hat{a}'\hat{a} \leqq k\hat{a}'B\hat{a}. \tag{3.1}$$

The theorem now follows from the lemma, equation (3.1), and the fact that $\hat{a}$ is Markoff relative to kB.

For the case $q = 1$, it can be shown that the minimax estimator is not unique, but it is not known whether it is unique for $q > 1$. However, the Markoff estimator $\hat{a}$ of Theorem 1 is the only linear minimax estimator, which can be seen as follows. A linear minimax estimator d must have bounded risk, and therefore must be unbiased. Suppose d is different from $\hat{a}$, and let $D = dd'/d'd$; then

$$kd'Dd = kd'd > ka'a,$$

i.e., the risk for d against the covariance kD is greater than the minimax risk.

Note that it follows immediately from Theorem 1, that if the characteristic roots of the covariance matrices in H are defined relative to any fixed symmetric positive definite matrix Q, then the Markoff estimator relative to Q will be minimax.

4. The case of bounds on the variances of the coordinates. In this section minimax estimators are found for the problem of Section 2 in the case in which the class H of covariances is delimited by bounding the variances of given linear functions of the random vector, in other words, by choosing a particular coordinate system and bounding the variances of the colrdinates.

THEOREM 2. *Let $k_1, \cdots, k_N$ be N given positive numbers; let H be the set of covariances C such that $c_{ii} \leqq k_i^2$ for $i = 1, \cdots, N$; then any $\hat{a}$ that minimizes $\sum_i k_i \mid a_i \mid$ subject to $T_a' = T'f$ (unbiasedness) is a minimax estimator for the problem of Section 2, and $\hat{c}^2 = (\sum_i k_i \mid a_i \mid)^2$ is the minimax loss.*

PROOF. There is no loss of generality in assuming that $k_i = 1$ for every i. As in Theorem 1, one is led to look for a least favorable covariance matrix among those of rank 1.

Let U be the set of linear unbiased estimators; for any C in H and b in U,

$$b'Cb = \sum_{ij} b_i b_j c_{ij} \leqq \sum_{ij} \mid b_i b_j \mid (c_{ii} c_{jj})^{\frac{1}{2}} \leqq \sum_{ij} \mid b_i b_j \mid = \left(\sum_i \mid b_i \mid\right)^2. \tag{4.1}$$

Let $\hat{a}$ be any vector that minimizes $\sum_i \mid a_i \mid$ in U, and let $\hat{c} = \sum_i \mid \hat{a}_i \mid$. By equation (4.1), and the lemma, the present theorem is proved if a vector $\hat{e}$ can be found such that (1) $\hat{a}$ is Markoff against $\hat{E} = \hat{e}\hat{e}'$; (2) $\hat{E}$ is in H, i.e., $\hat{e}_i^2 = 1$ for every i; and (3) the risk for $\hat{a}$ against E equals $\hat{c}^2$.

To this end, let S be the set of all vectors b such that $\sum_i \mid b_i \mid \leqq c$. S is a bounded convex polyhedron, and the intersection of S with U is contained in

the boundary of S, by the definition of c. Hence there is a hyperplane supporting S that contains U, i.e., there is a vector $\hat{e}$ such that

$$b'\hat{e} = \hat{c}, \text{ for all } b \text{ in } U, \tag{4.2}$$

$$b'\hat{e} \leqq \hat{c}, \text{ for all } b \text{ in } S \tag{4.3}$$

(see, for example, [2], p. 4).

By (4.2), e satisfies conditions (1) and (3) above. By the definition of S, any vector with one coordinate equal in absolute value to $\hat{c}$, and all other coordinates zero, is in S. Hence, by (4.3), $\hat{c} \mid \hat{e}_i \mid \leqq \hat{c}$, for every i, so that $\hat{e}_i^2 \leqq 1$ for every i; thus condition (2) above is also satisfied, which completes the proof.

Note that Theorem 2 characterizes all the linear minimax estimators, which is easily seen by an argument similar to that which follows Theorem 1.

5. Examples.

1. Suppose that the random variables $y_1, \cdots, y_N$ each have the same mean x, which is to be estimated, and assume that the sum of the variances of the y_i is not greater than k. To apply Theorem 1, Take T to be the $N \times 1$ matrix whose elements are all equal to 1, f to be vector for which $\sum f_i = 1$ (e.g., $[1, 0, \cdots, 0]$), and $q = 1$. It follows that a minimax estimate of $f'm_p = x$ is the arithmetic mean of $y_1, \cdots, y_N$, i.e., $\hat{a} = (1/N, \cdots, 1/N)$, and the minimax loss is $k \sum_i \hat{a}_i^2 = k/N$. This minimax estimator is, of course, the Markoff estimator for the situation in which it is known that the y_i are independent, with equal variances.

The same result would be obtained if it were assumed that the variance of any linear combination $\sum b_i y_i$ such that $\sum b_i^2 = 1$ is not greater than k (the case $q = \infty$).

2. Consider the estimation problem of Example 1, except now assume that the variance of y_i is not greater than k_i^2, $i = 1, \cdots, N$. By Theorem 2, a minimax estimator is given by

$$\hat{a}_i = \begin{cases} 1, & \text{for that } i \text{ for which } k_i \text{ is minimum,} \\ 0, & \text{otherwise,} \end{cases} \tag{5.1}$$

and the minimax loss is $\min_i k_i^2$. Note that in this example the minimax loss is independent of the sample size N, except insofar as $\min_i k_i$ depends upon N. If $k_1 = \cdots = k_N$, then any linear unbiased estimator is minimax.

3. Suppose it is required to estimate the slope e in the linear regression of one variable on another, and it is assumed that the variance of the "dependent variable" is not greater than k^2. To apply Theorem 2, take

$$T' = \begin{bmatrix} 1, \cdots, 1 \\ t_1, \cdots, t_N \end{bmatrix} \quad \text{and} \quad x' = (d, e),$$

where $t_1, \cdots, t_N$ are the values of the "independent variable," and d and e are unknown. A bounded risk (unbiased) linear estimator a must satisfy

$$\begin{aligned} \sum a_i &= 0, \\ \sum a_i t_i &= 1. \end{aligned} \tag{5.2}$$

By Theorem 2, any $\hat{a}$ that minimizes $\sum |a_i|$ subject to equation (5.2) is a minimax estimator of e. Without loss of generality, t_N can be taken to be the largest value of t_i, and t_1 the smallest; then it is not hard to show that the unique solution of the above minimization problem is

$$a_i = \begin{cases} \dfrac{-1}{t_N - t_1}, & \text{for} \quad i = 1, \\ \dfrac{1}{t_N - t_1}, & \text{for} \quad i = N, \\ 0 \quad , & \text{otherwise;} \end{cases} \tag{5.3}$$

and the minimax loss is $4k^2/(t_N - t_1)^2$. In other words, a minimax estimate of e is obtained by taking the slope of the line passing through the "extreme" points (y_1, t_1) and (y_N, t_N).

4. Consider the estimation problem of Example 3, but assume that the sum of the variances of $y_1, \cdots, y_N$ is not greater than k. As in Example 1, this corresponds to taking $q = 1$ in Theorem 1. By Theorem 1 the usual least squares estimate $\sum [(y_i - \bar{y})(t_i - \bar{t})]/(t_i - \bar{t})^2$ is a minimax estimate of e, and the minimax loss is $k/\sum (t_i - \bar{t})^2$.

Suppose further that $t_i = i - 1$ (e.g., think of t_i as successive times), and consider the transformation (taking successive differences)

$$z_i = \begin{cases} y_1, & \text{for } i = 1, \\ y_i - y_{i-1}, & \text{for } 1 = 2, \cdots, N. \end{cases} \tag{5.4}$$

The means of the z_i are

$$Ez_i = \begin{cases} d, & \text{for } i = 1, \\ e, & \text{for } i = 2, \cdots, N. \end{cases} \tag{5.5}$$

Now assume that the sum of the variances of the *new variables* z_i is not greater than k; then by Theorem 1 a minimax estimate of e is

$$\frac{1}{N-1} \sum_2^N z_i = \frac{y_N - y_1}{N - 1},$$

and the minimax loss is $k/(N - 1)$, a different result from that obtained before making the transformation (5.4).

REFERENCES

[1] A. C. AITKEN, "On least squares and the linear combination of observations," *Proc. Roy. Soc. Edinburgh*, Vol. 55 (1935).

[2] T. BONNESEN AND W. FENCHEL, *Theorie der Konvexen Körper*, J. Springer, Berlin, 1934.

[3] J. L. HODGES AND E. L. LEHMANN, "Some problems in minimax point estimation," *Ann. Math. Stat.*, Vol. 21 (1950), 182-197.

[4] L. J. SAVAGE, *The Foundations of Statistics*, John Wiley & Sons, New York, 1954.

COVARIANCES OF LEAST-SQUARES ESTIMATES WHEN RESIDUALS ARE CORRELATED[1]

By M. M. Siddiqui[2]

University of North Carolina

1. Summary. In this paper we will study the effects on the covariance matrix of the least-squares estimates of regression coefficients and on the estimate of the residual variance when the usual condition of independence of residuals is violated. The cases of linear trend and of regression on trigonometric functions will be considered in some detail.

2. Introduction. Several authors have studied the problem of estimating regression coefficients when residuals are autocorrelated. We refer here only to the work of Grenander and Rosenblatt [2, 3, 4]. Grenander [2] gives conditions on the regression variables for the existence of consistent estimates of the regression coefficients. He also gives conditions on the residual process under which the least-squares (L.S.) estimate of a regression coefficient is asymptotically efficient with respect to the Markov estimate. The covariances of the L.S. estimates as summarized in a matrix form are well known and are given at the end of section 3. The exact expression for an individual covariance or variance in the general case is easily extracted from this matrix and is given in section 4. The variance of the L.S. estimate in the general case is also given by Grenander [2, (8) p. 258]. Asymptotic expressions for the covariances of these estimates are also available [2, 4]. However, it seemed desirable to present here, in some detail, exact expressions or high order approximations to them for the individual variances and covariances of the L.S. estimates of regression coefficients and for the expectation of the estimate of residual variance, particularly for the cases of general interest, in readily usable form, and derived in an elementary fashion. The first term of each of our expressions coincide with the asymptotic expression given in [2, 4], when the regression coefficients are made comparable.

Bounds on the covariances of L.S. estimates are also provided in (7).

3. The L.S. estimates. Let $y = x'\beta + \Delta$ be the regression equation, where y and Δ are $N \times 1$ column vectors, β is a $p \times 1$ column vector, x is a $p \times N$ matrix and a prime is used to denote the transpose of a matrix or a vector. It is assumed that $N > p$, x is non-stochastic and of rank p, and Δ is a $N(0, \sigma^2 P)$ vector variate, where 0 is a zero vector and P is a positive definite correlation matrix.

Introducing $c = (xx')^{-1}$, $b = cxy$, $v = y - x'b$, $n = N - p$, and writing δq

---

Received December 3, 1957; revised June 13, 1958.

[1] Sponsored by the Office of Naval Research under the contract for research in probability and statistics at Chapel Hill. Reproduction in whole or in part is permitted for any purpose of the United States Government.

[2] Now at the Boulder Laboratories, National Bureau of Standards.

for $q - Eq$, where q is a variate with expected value Eq, it is known that b and $s^2 = v'v/n$ are the least-squares estimates of β and σ^2 respectively. It is also known that $Eb = \beta$, and

$$B = E\delta b\delta b' = \sigma^2 cxPx'c.$$

In case $P = I_N$, where I_N is the $N \times N$ identity matrix, $Es^2 = \sigma^2$ and $B = \sigma^2 c$.

4. The covariance matrix B. We propose to study the effects on B and Es^2 when P is given by

$$P = I_N + \sum_{k=1}^{N-1} \rho_k(C^k + C'^k), \tag{1}$$

where

$$C = \begin{bmatrix} 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & 0 & \cdots & 0 \end{bmatrix}, \tag{2}$$

i.e. when

$$E\Delta_t\Delta_{t\pm k} = \sigma^2\rho_k, \qquad k = 0, 1, \cdots, N-1, \rho_0 = 1. \tag{3}$$

We have

$$v = y - x'b = x'(\beta - b) + \Delta = (I_N - x'cx)\Delta \tag{4}$$

as $b = \beta + cx\Delta$. Writing $m = x'cx$, we have $m' = m$ and $m^2 = m$. Hence if λ is a characteristic root of m, $\lambda = 0$ or 1. Writing "tr" for the trace of a matrix we obtain tr $m = p$. Now, by simple evaluation

$$Es^2 = \frac{\sigma^2}{n}[N - \operatorname{tr} Pm] = \sigma^2\left[1 - \frac{2}{n}\sum_{k=1}^{N-1}\sum_{t=1}^{N-k} \rho_k\, m_{t,t+k}\right]. \tag{5}$$

Here, if e is a matrix, e_{ij} or $e_{i,j}$ refers to its element in the ith row and the jth column.

If we write $d = cx$, we find that

$$B_{ij} = E\delta b_i\,\delta b_j = \sigma^2\left[\sum_{t=1}^{N} d_{it}\,d_{jt} + \sum_{k=1}^{N-1}\sum_{t=1}^{N-k} \rho_k(d_{jt}\,d_{i,t+k} + d_{it}\,d_{j,t+k})\right].$$

If, by a proper choice of x or with a suitable transformation on x, we make $xx' = c^{-1} = I_p$, we have $d = x$. Writing x_i' for the row vector in the ith row of x, we find

$$B_{ij} - \sigma^2\delta_{ij} = \sigma^2\sum_{k=1}^{N-1} \rho_k\, x_i'(C^k + C'^k)x_j, \qquad i, j = 1, \cdots, p; \tag{6}$$

where $\delta_{ij} = 0$ if $i \neq j$ and $\delta_{ii} = 1$.

It has been shown [1, p. 130] that if A is an $N \times N$ real symmetric matrix with

characteristic roots $\alpha_1 \leqq \alpha_2 \leqq \cdots \leqq \alpha_N$, and u and v are $N \times 1$ real vectors, then, under the conditions $u'u = v'v = 1$, $u'v = 0$, the bilinear form $u'Ay$ has a maximum $(\alpha_N - \alpha_1)/2$ and a minimum $(\alpha_1 - \alpha_N)/2$. Also the quadratic form $u'Au \leqq \alpha_N$. Now the maximum characteristic root of $C^k + C'^k$, where k is a positive integer,

$$\alpha_N = 2\cos\left\{\frac{\pi}{\left[\frac{N+k-1}{k}\right]+1}\right\} \leqq 2\cos\left\{\frac{k\pi}{N+2k-1}\right\},$$

where $[q]$ denotes the largest integer $\leqq q$, and the minimum characteristic root $\alpha_1 = -\alpha_N$, [1, p. 101]. Hence, we obtain

$$(7) \qquad |B_{ij} - \sigma^2\delta_{ij}| < 2\sigma^2 \sum_{k=1}^{N-1} \left|\rho_k \cos\frac{k\pi}{N+2k-1}\right| < 2\sigma^2 \sum_{k=1}^{N-1} |\rho_k|.$$

In the case $\rho_k = \rho^k$, $k = 0, 1, \cdots$, where $\alpha = |\rho| < 1$, we have

$$(8) \qquad |B_{ij}| < \frac{2\alpha\sigma^2}{1-\alpha} \quad \text{if } i \neq j, \qquad B_{ii} < \sigma^2\left(\frac{1+\alpha}{1-\alpha}\right).$$

5. Linear trend. Let $N = 2r + 1$ where r is a positive integer and consider the linear trend in the form

$$(9) \qquad y_t = \beta_1(2r+1)^{-1/2} + \beta_2(t-r-1)/a + \Delta_t, \qquad t = 1, \cdots, N,$$

where

$$(10) \qquad a^2 = r(r+1)(2r+1)/3 = N(N^2-1)/12.$$

In the notation of section 3

$$(11) \qquad \begin{aligned} x_{1t} &= (2r+1)^{-1/2}, \qquad x_{2t} = (t-r-1)/a, \qquad t = 1, \cdots, N, \\ b_1 &= \sqrt{N}\bar{y}, \qquad b_2 = \left[\sum_{t=1}^{N} ty_t - (r+1)\sum_{t=1}^{N} y_t\right] \Big/ a. \end{aligned}$$

Furthermore

$$c = I_p, \qquad p = 2, \qquad n = N - 2,$$

$$m_{ij} = (x'x)_{ij} = \frac{1}{N} + \frac{3(2i-N-1)(2j-N-1)}{N(N^2-1)}$$

$$ns^2 = \sum y_t^2 - b_1^2 - b_2^2,$$

$$(12) \qquad B_{11} = \sigma^2\left[1 + 2\sum_{k=1}^{N-1}\left(1 - \frac{k}{N}\right)\rho_k\right], \qquad B_{12} = 0,$$

$$B_{22} = \sigma^2\left[1 + 2\sum_{k=1}^{N-1}\rho_k - \frac{2}{N}\left(3 + \frac{2}{N^2-1}\right)\sum_{k=1}^{N-1} k\rho_k + \frac{4}{N(N^2-1)}\sum_{k=1}^{N-1} k^3\rho_k\right]$$

and

$$Es^2 = \sigma^2\left[1 - \frac{4}{n}\sum_{k=1}^{N-1}\rho_k + \frac{2}{nN}\left(4 + \frac{2}{N^2-1}\right)\sum_{k=1}^{N-1}k\rho_k - \frac{4}{nN(N^2-1)}\sum_{k=1}^{N-1}k^3\rho_k\right]$$

For the case when $\rho_k = \rho^k$, we can evaluate the summations $\sum \rho_k$, $\sum k\rho_k$, etc. If N is moderately large we may neglect ρ^N and thus find

$$E\frac{s^2}{\sigma^2} \cong 1 - \frac{4\rho}{n(1-\rho)} + \frac{8\rho}{nN(1-\rho)^2} + \frac{4(\rho + 4\rho^2 + \rho^3)}{nN(N^2-1)(1-\rho)^4},$$

$$(13)\quad \begin{aligned} \frac{B_{11}}{\sigma^2} &= 1 - \frac{2}{N}\frac{\rho - N\rho^N + (N-1)\rho^{N+1}}{(1-\rho)^2} \\ &\quad + \frac{2(\rho - \rho^N)}{1-\rho} \cong \frac{1+\rho}{1-\rho} - \frac{2\rho}{N(1-\rho)^2}, \\ B_{12} &= 0, \qquad \frac{B_{22}}{\sigma^2} \cong \frac{1+\rho}{1-\rho} - \frac{6\rho}{N(1-\rho)^2}. \end{aligned}$$

We note that b_i are independently distributed $N(\beta_i, B_{ii})$, $i = 1, 2$, variates. If we set

$$b_1' = N^{-1/2}b_1 = \bar{y}, \qquad b_2' = \frac{\sqrt{12}b_2}{\sqrt{N(N^2-1)}},$$

the estimate of Ey_t is given by

$$(14)\quad Y_t = \bar{y} + b_2'(t - r - 1)$$

and under the first order autoregressive scheme for Δ's,

$$(15)\quad \sigma_{\bar{y}}^2 \cong \frac{\sigma^2}{N}\left(\frac{1+\rho}{1-\rho}\right), \qquad \sigma_{b_2'}^2 \cong \frac{12\sigma^2}{N(N^2-1)}\left(\frac{1+\rho}{1-\rho}\right), \qquad \operatorname{cov}(\bar{y}, b_2') = 0.$$

Thus

$$\sigma_{Y_t}^2 \cong \frac{\sigma^2}{N}\left(\frac{1+\rho}{1-\rho}\right)\left[1 + \frac{12(t-r-1)^2}{N^2-1}\right].$$

6. Regression on trigonometric functions. Consider

$$(16)\quad \begin{aligned} y_t &= \beta_1/\sqrt{N} + \sqrt{2/N}\sum_{i=1}^{q}\beta_{2i}\cos\lambda_i t \\ &\quad + \sqrt{2/N}\sum_{i=1}^{q}\beta_{2i+1}\sin\lambda_i t + \Delta_t, \qquad t = 1, \cdots, N, \end{aligned}$$

where $\lambda_i = 2\pi\omega_i/N$ and ω_i is a positive integer less than N for $i = 1, 2, \cdots, q$ and $\omega_i \neq \omega_j$ if $i \neq j$.

In the notation of section 2

$$x_{1t} = 1/\sqrt{N}, \qquad x_{2i,t} = \sqrt{2/N}\cos\lambda_i t,$$
$$x_{2i+1,t} = \sqrt{2/N}\sin\lambda_i t,\ i = 1, 2, \cdots, q;\ t = 1, 2, \cdots, N,$$
$$xx' = c^{-1} = I_{2q+1}, \qquad n = N - 2q - 1,$$

(17)
$$b_1 = \sqrt{N}\bar{y}, \qquad b_{2i} = \sqrt{2/N}\sum_t y_t \cos\lambda_i t,$$
$$b_{2i+1} = \sqrt{2/N}\sum_t y_t \sin\lambda_i t,\ i = 1, \cdots, q,$$
$$m_{ts} = 1/N + 2/N\sum_{i=1}^{q}\cos(t-s)\lambda_i, \qquad t, s = 1, \cdots, N,$$
$$s^2 = \left(\sum_t y_t^2 - \sum_{i=1}^{2q+1} b_i^2\right)\Big/ n.$$

We find

$$E\frac{s^2}{\sigma^2} = 1 - \frac{2}{n}\sum_{k=1}^{N-1}\rho_k + \frac{2}{nN}\sum_{k=1}^{N-1}k\rho_k - \frac{4}{n}\sum_{k=1}^{N-1}\sum_{i=1}^{q}\left(1 - \frac{k}{N}\right)\rho_k\cos k\lambda_i. \tag{18}$$

For the covariances of b_i and b_j we obtain

(19)
$$B_{11} = \sigma^2\left[1 + 2\sum_{k=1}^{N-1}\left(1 - \frac{k}{N}\right)\rho_k\right],$$
$$B_{1,2i} = \frac{\sqrt{2}\,\sigma^2}{N}\sum_{k=1}^{N-1}\sum_{t=1}^{N-k}\rho_k\{\cos(t+k)\lambda_i + \cos t\lambda_i\}, \qquad i = 1, \cdots, q,$$
$$B_{2i,2i} = \sigma^2\left[1 + \frac{4}{N}\sum_{k=1}^{N-1}\sum_{t=1}^{N-k}\rho_k\cos(k+t)\lambda_i\cos t\lambda_i\right], \qquad i = 1, \cdots, q.$$
$$B_{2i,2j+1} = \frac{2\sigma^2}{N}\sum_{k=1}^{N-1}\sum_{t=1}^{N-k}\rho_k\{\cos t\lambda_i\sin(k+t)\lambda_j + \cos(k+t)\lambda_i\sin t\lambda_j\}, \qquad i, j = 1, \cdots, q.$$

$B_{1,2i+1}$ and $B_{2i+1,2i+1}$ are obtainable from the expressions for $B_{1,2i}$ and $B_{2i,2i}$ respectively by replacing cosine by sine.

If $\rho_k = \rho^k$, and ρ^N is negligible, we find for the variances, after some reduction,

(20)
$$\frac{B_{11}}{\sigma^2} \cong \frac{1+\rho}{1-\rho} - \frac{2\rho}{N(1-\rho)^2},$$
$$\frac{B_{2i,2i}}{\sigma^2} \cong \frac{1-\rho^2}{1 - 2\rho\cos\lambda_i + \rho^2} + \frac{\rho\cos\lambda_i}{N(1 - 2\rho\cos\lambda_i + \rho^2)} - \frac{\rho(1+\rho^2)\cos\lambda_i - 2\rho^2}{N(1 - 2\rho\cos\lambda_i + \rho^2)^2},$$
$$\frac{B_{2i+1,2i+1}}{\sigma^2} \cong \frac{1-\rho^2}{1 - 2\rho\cos\lambda_i + \rho^2} - \frac{\rho\cos\lambda_i}{N(1 - 2\rho\cos\lambda_i + \rho^2)} - \frac{\rho(1+\rho^2)\cos\lambda_i - 2\rho^2}{N(1 - 2\rho\cos\lambda_i + \rho^2)^2}, \qquad i = 1, \cdots, q.$$

Also

$$E\frac{s^2}{\sigma^2} \cong 1 - \frac{2\rho}{n(1-\rho)} - \frac{4\rho}{n}\sum_{i=1}^{q}\frac{\cos\lambda_i - \rho}{1 - 2\rho\cos\lambda_i + \rho^2} + O\left(\frac{1}{N^2}\right). \tag{21}$$

7. Concluding remarks. We conclude with the remarks that in most practical cases the correlation matrix for Δ's will not be known. However, if Δ's may be represented as a stationary autoregressive process of some small order—in many cases first or second order scheme gives a reasonably good fit—we would be required to estimate a few parameters $\rho_1, \rho_2, \cdots, \rho_k$. We, however, note that these quantities do not appear in b and s^2, only in B and Es^2.

We further observe that the estimates, $\hat{\beta}$ and $\hat{\sigma}^2$, of β and σ^2 obtained from maximizing the likelihood function will depend on the parameters of P, i.e. on $\rho_1, \rho_2, \cdots, \rho_{N-1}$, which will mean using sample serial correlation coefficients to estimate ρ's in the expression for $\hat{\beta}$ and $\hat{\sigma}^2$. These estimates will obviously be non-linear. Thus it seems more desirable to stick to the least-squares estimates b and s^2 rather than to attempt to develop maximum-likelihood (or minimum χ^2) estimates.

8. Acknowledgement. The writer wishes to express his indebtedness to Professor Harold Hotelling for drawing his attention to this problem.

REFERENCES

[1] M. M. SIDDIQUI, *Distributions of Some Serial Correlation Coefficients*, University of North Carolina Institute of Mathematical Statistics Mimeograph Series No. 164, Chapel Hill (1957).

[2] ULF GRENANDER, "On the estimation of regression coefficients in the case of an autocorrelated disturbance," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 252–272.

[3] MURRAY ROSENBLATT, "On the estimation of regression coefficients of a vector-valued time series with a stationary residual," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 99–121.

[4] ULF GRENANDER AND MURRAY ROSENBLATT, *Statistical Analysis of Stationary Time Series*, John Wiley and Sons, New York, 1957, Chapter 7.

ON A PROBABILITY PROBLEM IN THE THEORY OF COUNTERS

By L. Takács

Research Institute for Mathematics, Hungarian Academy of Sciences, Budapest, Hungary

1. Introduction. Let us suppose that particles arrive at a counter in the time interval $(0, \infty)$ according to a Poisson-process of density λ. Each particle arriving in the time interval $(0, \infty)$ independently of the others gives rise to an impulse with probability p or 1 according to whether at this instant there is an impulse present or there is no impulse present. The time durations of the impulses are identically distributed independent positive random variables with distribution function $H(x)$ and these random variables are independent of the instants of the arrivals and of the events of the realizations of the impulses. We define as "registered particles" those particles which occur at an instant when there is no impulse present. Denote by ν_t the number of the registered particles in the time interval $(0, t)$. The problem is to determine the distribution law of ν_t and its asymptotic behaviour as $t \to \infty$.

The particular case of the above problem, when the time durations of the impulses are constant, was investigated earlier by G. E. Albert and L. Nelson [1].

2. The structure of the process. Denote by $\{\tau_n\}$ the sequence of instants at which particles are registered. We say that the system at any instant t is in state A when no impulse covers the instant t and in state B otherwise. Then the system assumes the states $A, B, A, B, \cdots$ alternatingly. Let us denote by ξ_1, η_1, ξ_2, η_2, $\cdots$ the times spent in states A and B respectively. If the system at the instant t is in state A, then t is evidently a regeneration point of the process. Consequently $\{\xi_n\}$ and $\{\eta_n\}$ are independent sequences of identically distributed positive random variables. Clearly $\mathbf{P}\{\xi_n \leq x\} = F(x) = 1 - e^{-\lambda x}$ if $x \geq 0$. Write $\mathbf{P}\{\eta_n \leq x\} = U(x)$, where $U(x)$ is still unknown. (We use $\mathbf{P}$ for the symbol of probability and $\mathbf{E}$ for expectation.) It can easily be seen that the instants of the transitions $A \to B$ coincide with the instants τ_n $(n = 1, 2, \cdots)$. Consequently the time differences $\tau_{n+1} - \tau_n$ $(n = 1, 2, \cdots)$ are identically distributed independent random variables with distribution function $G(x) = F(x) * U(x)$ i.e.

$$G(x) = \int_0^x U(x-y)e^{-\lambda y}\lambda\, dy, \tag{1}$$

while $\mathbf{P}\{\tau_1 \leq x\} = F(x)$.

3. Notations. Let us introduce the following Laplace-Stieltjes transforms:

$$\gamma(s) = \int_0^\infty e^{-sx}\, dG(x) \tag{2}$$

Received May 17, 1957; revised May 15, 1958.

and

$$\omega(s) = \int_0^\infty e^{-sx}\, dU(x). \tag{3}$$

By (1) we have

$$\gamma(s) = \frac{\lambda}{\lambda + s}\,\omega(s). \tag{4}$$

Further put

$$\alpha = \int_0^\infty x\, dH(x), \qquad \beta^2 = \int_0^\infty (x-\alpha)^2\, dH(x), \tag{5}$$

$$\tau = \int_0^\infty x\, dU(x), \qquad \rho^2 = \int_0^\infty (x-\tau)^2\, dU(x), \tag{6}$$

$$A = \int_0^\infty x\, dG(x), \qquad B^2 = \int_0^\infty (x-A)^2\, dG(x). \tag{7}$$

By (1) we clearly have that $A = \tau + (1/\lambda)$ and $B^2 = \rho^2 + (1/\lambda^2)$.

Denote by $P(t)$ the probability that at the instant t the system is in state A, and put

$$\pi(s) = \int_0^\infty e^{-st} P(t)\, dt. \tag{8}$$

4. Theorems concerning ν_t. In what follows we shall give some general theorems for ν_t.

1. We have

$$\mathbf{P}\{\nu_t \leqq n\} = 1 - F(t) * G_n(t), \tag{9}$$

where $G_n(x)$ denotes the n-fold convolution of $G(x)$ with itself. ($G_0(x) = 1$ if $x \geqq 0$ and $G_0(x) = 0$ if $x < 0$). For

$$\mathbf{P}\{\nu_t \leqq n\} = \mathbf{P}\{t < \tau_{n+1}\} = 1 - \mathbf{P}\{\tau_{n+1} \leqq t\},$$

and $\tau_{n+1} = \tau_1 + (\tau_2 - \tau_1) + \cdots + (\tau_{n+1} - \tau_n)$ is a sum of independent random variables.

2. If $A < \infty$, then we have

$$\lim_{T\to\infty} \mathbf{P}\{\nu_{T+t} - \nu_T \leqq n\} = 1 - G^*(t) * G_n(t), \tag{10}$$

where

$$G^*(t) = \begin{cases} \dfrac{1}{A}\displaystyle\int_0^t [1 - G(u)]\, du & \text{if } \ t \geqq 0 \\ 0 & \text{if } \ t < 0. \end{cases}$$

The proof is similar to that of (9), only we must use the result

$$\lim_{T\to\infty} \mathbf{P}\{\nu_{T+t} - \nu_T \geqq 1\} = G^*(t),$$

which was proved by J. L. Doob [2].

3. If $B^2 < \infty$, then we have

$$\lim_{t\to\infty} \mathbf{P}\left\{\frac{\nu_t - \dfrac{t}{A}}{\sqrt{\dfrac{B^2 t}{A^3}}} \leqq x\right\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-u^2/2}\, du. \tag{11}$$

This can be proved by the aid of the method of W. Feller [3]. (Cf. [5]).

4. If $B^2 < \infty$, then we have

$$\mathbf{P}\left\{\limsup_{t\to\infty} \frac{\nu_t - \dfrac{t}{A}}{\sqrt{\dfrac{2B^2}{A^3} t \log\log t}} = 1\right\} = \mathbf{P}\left\{\liminf_{t\to\infty} \frac{\nu_t - \dfrac{t}{A}}{\sqrt{\dfrac{2B^2}{A^3} t \log\log t}} = -1\right\} = 1. \tag{12}$$

This can be proved by the aid of the law of the iterated logarithm stated by P. Hartman and A. Wintner [4].

5. Applying the strong law of large numbers we obtain

$$\mathbf{P}\left\{\lim_{t\to\infty} \frac{\nu_t}{t} = \frac{1}{A}\right\} = 1, \tag{13}$$

(cf. J. L. Doob [2]).

It is easy to see that $\mathbf{E}\{\nu_t\} = M(t)$ can be expressed as follows:

$$M(t) = \sum_{n=1}^{\infty} \mathbf{P}\{\tau_n \leqq t\}. \tag{14}$$

6. If $A < \infty$, then for any $h > 0$ we have

$$\lim_{t\to\infty} \frac{M(t+h) - M(t)}{h} = \frac{1}{A}, \tag{15}$$

by the theorem of J. L. Doob [2].

7. If $B^2 < \infty$, then we have

$$\int_0^\infty e^{-st}\, dM(t) = \frac{1}{As} + \frac{B^2 + A^2}{2A^2} - \frac{1}{\lambda A} + o(s) \tag{16}$$

if $s \to 0$. For by (9) and (14) we have

$$\int_0^\infty e^{-st}\, dM(t) = \frac{\lambda}{(\lambda + s)[1 - \gamma(s)]} \tag{17}$$

and

$$\gamma(s) = 1 - sA + \frac{s^2}{2}(B^2 + A^2) + o(s^2)$$

if $s \to 0$.

8. For the Laplace-transform of $P(t)$ we have

$$\pi(s) = \int_0^\infty e^{-st} P(t)\, dt = \frac{1}{(\lambda + s)[1 - \gamma(s)]}, \tag{18}$$

and

$$P = \lim_{t\to\infty} P(t) = \frac{1}{\lambda A}. \tag{19}$$

PROOF. As $M(t + \Delta t) = M(t) + P(t)\lambda\Delta t + o(\Delta t)$, we have $M'(t) = \lambda P(t)$, and thus (18) follows from (17). Now

$$P(t) = 1 - \int_0^t [1 - U(t - x)]\, dM(x), \tag{20}$$

for by the theorem of total probability we have

$$1 - P(t) = \sum_{n=1}^\infty \int_0^t [1 - U(t - x)]\, d\mathbf{P}\{\tau_n \leqq x\} = \int_0^t [1 - U(t - x)]\, dM(x),$$

which agrees with (20). By virtue of (15) we obtain from (20)

$$\lim_{t\to\infty} P(t) = 1 - \frac{1}{A}\int_0^\infty [1 - U(x)]\, dx = 1 - \frac{\tau}{A}.$$

Since $\tau = A - (1/\lambda)$, equation (19) follows.

REMARK. Taking into consideration that $M'(t) = \lambda P(t)$, we obtain from (20) the following integral equation for $P(t)$:

$$P(t) = 1 - \lambda \int_0^t [1 - U(t - x)]P(x)\, dx. \tag{21}$$

From (18) or from (21) we obtain that

$$\omega(s) = \int_0^\infty e^{-sx}\, dU(x) = \frac{\lambda + s}{\lambda}\left[1 - \frac{1}{(\lambda + s)\pi(s)}\right]. \tag{22}$$

To apply the above theorems it remains only to determine $G(x)$, A, and B^2.

5. The determination of $G(x)$, A, and B^2.

THEOREM. *If $0 < p \leqq 1$, then we have*

$$\gamma(s) = \int_0^\infty e^{-sx}\, dG(x) = \frac{\lambda p + s}{p(\lambda + s)} - \frac{1}{p(\lambda + s)}\left\{\int_0^\infty \exp\left[-st - \lambda p \int_0^t (1 - H(x))\, dx\right] dt\right\}^{-1}; \tag{23}$$

if $\alpha < \infty$, then

$$A = \frac{e^{\lambda p \alpha} + p - 1}{\lambda p}, \tag{24}$$

and if $\beta^2 < \infty$, then

$$B^2 = \frac{2e^{\lambda p \alpha}}{\lambda p} \int_0^\infty \left\{\exp\left[\lambda p \int_t^\infty (1 - H(x))\, dx\right] - 1\right\} dt + \frac{2e^{\lambda p \alpha} - e^{2\lambda p \alpha} + p^2 - 1}{(\lambda p)^2}. \tag{25}$$

If $p = 0$ then $U(x) = H(x)$ and consequently

$$G(x) = \int_0^x H(x - y) e^{-\lambda y} \lambda\, dy, \tag{26}$$

$$A = \frac{1 + \lambda\alpha}{\lambda}, \tag{27}$$

and

$$B^2 = \frac{1 + \lambda^2\beta^2}{\lambda^2}. \tag{28}$$

PROOF. Let us consider a new process which is a particular case of the process defined in the Introduction. Suppose that the density of the underlying Poisson-process is λ^* and each particle gives rise to an impulse (with probability $p^* = 1$). Let $H^*(x) = H(x)$ be the distribution function of the duration of the impulses. This is the case of Type II counter. Denote by $\{\xi_n^*\}$ and $\{\eta_n^*\}$ the sequences of the times spent in state A and B respectively. Clearly $\mathbf{P}\{\xi_n^* \leqq x\} = 1 - e^{-\lambda^* x}$ if $x \geqq 0$. Write $\mathbf{P}\{\eta_n^* \leqq x\} = U^*(x)$. Denote by $P^*(t)$ the probability that at the instant t there is no impulse present. We have showed in [5] that

$$P^*(t) = \exp\left[-\lambda^* \int_0^t [1 - H(x)]\, dx\right]. \tag{29}$$

Applying (22) it follows that

$$\omega^*(s) = \int_0^\infty e^{-sx}\, dU^*(x) = \frac{\lambda^* + s}{\lambda^*} - \frac{1}{\lambda^*\pi^*(s)}, \tag{30}$$

where

$$\pi^*(s) = \int_0^\infty e^{-st} P^*(t)\, dt = \int_0^\infty \exp\left\{-st - \lambda^* \int_0^t [1 - H(x)]\, dx\right\} dt. \tag{31}$$

Now we observe that, if in this new process $\lambda^* = \lambda p$, then we have

$$U^*(x) = U(x), \tag{32}$$

where $U(x)$ is related to the general process. The equality (32) can easily be seen if we take into consideration that the arrivals of those particles in the general process, which arrive during a dead time and which give rise to an impulse, form a Poisson process with density λp. Accordingly by (30) and (31) we have

$$\begin{aligned}\omega(s) = \int_0^\infty e^{-sx}\, dU(x) = \frac{\lambda p + s}{\lambda p} \\ - \left\{\lambda p \int_0^\infty \exp\left[- st - \lambda p \int_0^t (1 - H(x))\, dx\right] dt\right\}^{-1},\end{aligned} \tag{33}$$

and by (4) we obtain (23), which was to be proved.

If we introduce for the new process the analogous quantities $M^*(t)$, A^* and B^{*2} corresponding to (7) and (14), then by (16) we obtain that

$$\int_0^\infty e^{-st}\, dM^*(t) = \lambda^* \int_0^\infty e^{-st}\, P^*(t)\, dt = \frac{1}{A^*s} + \frac{B^{*2} + A^{*2}}{2A^{*2}} - \frac{1}{\lambda^* A^*} + o(s) \tag{34}$$

if $s \to 0$. Since $P^* = \lim_{t\to\infty} P^*(t) = e^{-\lambda^* \alpha}$, we obtain from (34) that

$$A^* = e^{\lambda^*\alpha} / \lambda^* \tag{35}$$

and, further,

$$B^{*2} = 2\lambda^* A^{*2} \int_0^\infty [P^*(t) - P^*]\, dt - A^{*2} + 2A^*/\lambda^*. \tag{36}$$

If in particular $\lambda^* = \lambda p$, then clearly

$$A - \frac{1}{\lambda} = A^* - \frac{1}{\lambda^*}$$

and

$$B^2 - \frac{1}{\lambda^2} = B^{*2} - \frac{1}{\lambda^{*2}},$$

and thus (24) and (26) are proved. The case $p = 0$ is evident.

Finally we remark that the more general case when the arrivals of the par-

ticles to the counter form a recurrent process was dealt by the author [6], [7], [8], but explicit solution was given only for a particular distribution $H(x)$.

REFERENCES

[1] G. E. Albert and L. Nelson, "Contributions to the statistical theory of counters data, *Ann. Math. Stat.*, 24 (1953), 9-22.

[2] J. L. Doob, "Renewal theory from the point of view of the theory of probability," *Trans. Amer. Math. Soc.*, 63 (1948), pp. 422-438.

[3] W. Feller, "Fluctuation theory of recurrent events," *Trans. Amer. Math. Soc.*, 67 (1949), pp. 98-119.

[4] P. Hartman and W. Wintner, "On the law of the iterated logarithm," *Amer. J. Math.*, 63 (1941), pp. 169-176.

[5] L. Takács, "On a probability problem arising in the theory of counters," *Proc. Cambr. Phil. Soc.*, 52 (1956), pp. 488-498.

[6] L. Takács, "On stochastic processes in the theory of counters," (Hungarian), *Magy. Tud. Akad.* III. *Oszt. Közl.*, 6 (1956), pp. 369-421.

[7] L. Takács, "On the sequence of events, selected by a counter from a recurrent process of events," *The Theory of Probability and Its Applications* (Moscow) 1 (1956), pp. 90-102.

[8] L. Takács, "On some probability problems concerning the theory of counters," *Acta Math. Acad. Sci. Hung.*, 8 (1957), pp. 127-138.

NOTES

DISTRIBUTION OF LINEAR CONTRASTS OF ORDER STATISTICS[1]

By Jacques St-Pierre

University of Montreal

Introduction. Many theoretical and practical problems of statistical nature have lead investigators to study methods capable of pooling the information contained in the ordered (or ranked) sample values with some properties of the assumed distribution of the parent population. Since, in analysis of variance situations, contrasts between functions of observations are of utmost importance, linear contrasts of order statistics will be considered here under the assumption that the underlying distribution is normal.

Null distribution of linear contrasts of order statistics. Let $x_0, x_1, \cdots, x_n$ denote $n + 1$ independent normal random variables with unknown means $\mu_1, \mu_2, \cdots, \mu_n$ respectively, and with a common variance $\sigma^2 = 1$ (say). Let $x_{(0)} > x_{(1)} > \cdots > x_{(n)}$ be the ordered values. Consider the following linear contrast

$$z = x_{(0)} - c_1 x_{(1)} - c_2 x_{(2)} - \cdots - c_n x_{(n)}, \qquad \sum_{i=1}^{n} c_i = 1;$$

$$0 \leqq c_i \leqq 1, \quad i = 1, \cdots, n.$$

Using, as a starting point, the joint density of $x_{(0)}, x_{(1)}, \cdots, x_{(n)}$ as given by Wilks [7], and with the help of appropriate transformations, the null distribution of z can be obtained. It takes the form of a rather messy expression containing a n-fold iterated integral. An interesting particular case: the density of the difference between the two largest ordered values can be obtained from the general form. St-Pierre and Zinger [6] have tabulated the null density of $u = x_{(0)} - x_{(1)}$ using a slightly different method.

It is of interest to consider the above contrast in the case of three random variables. The density of $z = x_{(0)} - cx_{(1)} - (1 - c)x_{(2)}$, under the hypothesis H_0: $\mu_0 = \mu_1 = \mu_2 = 0$ (say), takes the form

$$g(z) = 3[\pi(c^2 - c + 1)]^{-1/2} \exp[-z^2/4(c^2 - c + 1)] \cdot \int_{(2c-1)z/[6(c^2-c+1)]^{1/2}}^{(c+1)z/(1-c)[6(c^2-c+1)]^{1/2}} (2\pi)^{-1/2} \exp(-t^2/2)\, dt. \tag{1}$$

With the help of [3], [4], and [5], $g(z)$ can be tabulated. Values of $g(z)$ are given in Table I for several values of the parameter c.

From the general form (1), several densities can be derived as particular cases. For instance, the value $c = 0$ leads to the density of the range as given by McKay

Received March 28, 1955; revised July 7, 1957.

[1] Part of a Ph.D. thesis presented to the Institute of Statistics, University of North Carolina. This research was supported in part by the United States Air Force, through the Office of Scientific Research of the Air Research and Development Command.

TABLE I

Values of $g(z)$, for various values of the constant c, where $z = x_{(0)} - cx_{(1)} - (1-c)x_{(2)}$

z \ c	0	0.1	0.2	0.4	0.6	0.8	0.9	1.0
0.0	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.84628
0.2	.10917	.12101	.13709	.17898	.26659	.49282	.73839	.78334
0.4	.21095	.23318	.26050	.33877	.47562	.70969	.75554	.70763
0.6	.29932	.32877	.36410	.45941	.59859	.71048	.67281	.62378
0.8	.36927	.40194	.43988	.53834	.64049	.63299	.58344	.53652
1.0	.41774	.44958	.48459	.55897	.60386	.54116	.49344	.45022
1.2	.44376	.47102	.49861	.54388	.53555	.45020	.40687	.36855
1.4	.44833	.46822	.48548	.49838	.45187	.36497	.32709	.29429
1.6	.43408	.44502	.45086	.43473	.36725	.28832	.25636	.22920
1.8	.40476	.40647	.40149	.36270	.28937	.22196	.19588	.17410
2.0	.36474	.35800	.34410	.29160	.22168	.16649	.14590	.12896
2.2	.31842	.30485	.28468	.22659	.16529	.12170	.10593	.09315
2.4	.26981	.25145	.23115	.17064	.12000	.08668	.07497	.06560
2.6	.22221	.20121	.17665	.12479	.08484	.06016	.05171	.04504
2.8	.17809	.15639	.13290	.08871	.05840	.04068	.03477	.03016
3.0	.13903	.11819	.09715	.06135	.03918	.02679	.02278	.01968
3.2	.10580	.08692	.06904	.04130	.02556	.01720	.01455	.01252
3.4	.07853	.06225	.04775	.02707	.01625	.01076	.00905	.00777
3.6	.05690	.04345	.03215	.01727	.01006	.00656	.00548	.00469
3.8	.04026	.02975	.02109	.01073	.00606	.00389	.00327	.00277
4.0	.02782	.01971	.01348	.00650	.00356	.00225	.00188	.00159

and Pearson [2]; while the value $c = 0.5$ leads to the density of $v = x_{(0)} - (x_{(0)} + x_{(1)} + x_{(2)})/3$ as given by McKay [1]. The complexity of the expression for $g(z)$ increases rapidly with the number of variables; consequently, we will limit our presentation to the above mentioned case.

Non-null distribution of linear contrasts of order statistics. Here again, and for the same reasons, only the case of three variables will be presented. In order to get the non-null distribution of $z = x_{(0)} - cx_{(1)} - (1-c)x_{(2)}$ the joint density of $x_{(0)}$, $x_{(1)}$ and $x_{(2)}$ must be used as a starting point. It is of the form

$$g(x_{(0)}, x_{(1)}, x_{(2)}) = \frac{1}{(2\pi)^{3/2}} \exp\left[\frac{-\mu'\mu}{2}\right] \exp\left[\frac{-X'X}{2}\right] \sum{}^{*} \exp(\mu_i' X),$$

where

$$\mu = \begin{vmatrix} \mu_0 \\ \mu_1 \\ \mu_2 \end{vmatrix}, \qquad X = \begin{vmatrix} x_0 \\ x_1 \\ x_2 \end{vmatrix}, \qquad \mu_i = \begin{vmatrix} \mu_{i_0} \\ \mu_{i_1} \\ \mu_{i_2} \end{vmatrix}$$

and $\sum^*$ stands for the summation over all the permutations i_0, i_1, i_2 of the numbers 0, 1 and 2. Introducing the contrast z with the appropriate transformation and integregating out the extra variables, one gets, after a few simplifications,

the following expression for the non-null density of z:

$$
\begin{aligned}
f(z) &= \frac{1}{\sqrt{2\pi}\,\sqrt{2(c^2 - c + 1)}} \exp\left[-\frac{1}{2}(\mu'\mu - M^2/3)\right] \\
&\cdot \sum{}^{*} \left\{ \exp\left[\frac{-(z^2 - 2\gamma_1 z)}{4(c^2 - c + 1)}\right] \exp\left[\frac{(\gamma_1 + 2\gamma_2)^2}{12(c^2 - c + 1)}\right] \right. \\
&\left. \cdot \int_{[(2c-1)z-(\gamma_1+2\gamma_2)]/[6(c^2-c+1)]^{1/2}}^{[(c+1)z-(1-c)(\gamma_1+2\gamma_2)]/[(1-c)(6(c^2-c+1))^{1/2}]} (2\pi)^{-1/2} \exp(-t^2/2)\, dt \right\},
\end{aligned} \tag{2}
$$

where $\gamma_1 = \mu_{i_0} - c\mu_{i_1} - (1 - c)\mu_{i_2}$, $\gamma_2 = -(1 - c)\mu_{i_0} + \mu_{i_1} - c\mu_{i_2}$, and $M = \mu_0 + \mu_1 + \mu_2$. It is easy to see, looking at (2), how much more complicated an expression for $f(z)$ can become in the case of several variables.

Many particular cases of interest have been considered, using expression (2) as a starting point. Only two cases are reported here. The first one corresponds to the hypothesis $H_1\colon \mu_0 = \delta$, $\mu_1 = \mu_2 = 0$, $\delta > 0$. Denoting by $f(z \mid H_1)$ the density of z under the hypothesis H_1, one gets

$$
f(z \mid H_1) = \frac{1}{\sqrt{\pi(c^2 - c + 1)}} \exp\left[(-\delta^2/3)(g_1 + g_2 + g_3)\right],
$$

where g_1, g_2 and g_3 are functions of z and of the parameters δ and c given by

$$
g_1(z; \delta, c) = \exp\left[-(3z^2 - 6\delta z - (2c - 1)^2\delta^2)/12(c^2 - c + 1)\right] I_1(z; \delta, c),
$$

$$
g_2(z; \delta, c) = \exp\left[-(3z^2 + 6c\delta z - (2 - c)^2\delta^2)/12(c^2 - c + 1)\right] I_2(z; \delta, c),
$$

$$
g_3(z; \delta, c) = \exp\left[-(3z^2 + 6(1 - c)\delta z - (1 + c)^2\delta^2)/12(c^2 - c + 1)\right] I_3(z; \delta, c).
$$

The functions I_1, I_2, and I_3 are given by

$$
I_1 = \int_{\frac{(2c-1)z-(2c-1)\delta}{[6(c^2-c+1)]^{1/2}}}^{\frac{[(c+1)z-(1-c)(2c-1)\delta]}{(1-c)[6(c^2-c+1)]^{1/2}}} \frac{\exp(-t^2/2)}{(2\pi)^{1/2}}\, dt, \qquad I_2 = \int_{\frac{(2c-1)z-(2-c)\delta}{[6(c^2-c+1)]^{1/2}}}^{\frac{(c+1)z-(1-c)(2-c)\delta}{(1-c)[6(c^2-c+1)]^{1/2}}} \frac{\exp(-t^2/2)}{(2\pi)^{1/2}}\, dt
$$

$$
I_3 = \int_{\frac{(2c-1)z+(1+c)\delta}{[6(c^2-c+1)]^{1/2}}}^{\frac{(c+1)z+(1+c)(1-c)\delta}{(1-c)[6(c^2-c+1)]^{1/2}}} \frac{\exp(-t^2/2)}{(2\pi)^{1/2}}\, dt.
$$

Table II contains the values of $f(z \mid H_1)$, in the case $\delta = 1$, for several values of the parameter c.

The case of equal spacing of the true means, i.e., the one corresponding to the hypothesis $H_2\colon \mu_0 = 2\delta$, $\mu_1 = \delta$, $\mu_2 = 0$, yields a slightly more complicated expression for $f(z \mid H_2)$. Table III contains some values of $f(z \mid H_2)$, in the particular case $\delta = 1$, for a few values of the parameter c.

TABLE II

Values of the density of $z = x_{(0)} - cx_{(1)} - (1 - c)x_{(2)}$ *under the hypothesis* $H_1 : \mu_0 = \delta = 1;\ \mu_1 = \mu_2 = 0$

c	0	0.1	0.2	0.4	0.6	0.8	0.9	1.0
0.0	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.69550
0.2	.07843	.08707	.09783	.12988	.19255	.36249	.58377	.65223
0.4	.15340	.16984	.19015	.24916	.35644	.56737	.63842	.60265
0.6	.22169	.24434	.27187	.34847	.47043	.60539	.58628	.54862
0.8	.28049	.30717	.33879	.42129	.52678	.56566	.52858	.49195
1.0	.32764	.35584	.38797	.46387	.53322	.50692	.46915	.43436
1.2	.36168	.38868	.41801	.47716	.51027	.44557	.40985	.37744
1.4	.38200	.40550	.42904	.46589	.45370	.38510	.35211	.32259
1.6	.38882	.40689	.42265	.43486	.39556	.32714	.29733	.27098
1.8	.38314	.39449	.40156	.39238	.33636	.27293	.24660	.22357
2.0	.36658	.37079	.36936	.34034	.27999	.22364	.20067	.18002
2.2	.34126	.33851	.32948	.28770	.22839	.17958	.16014	.14373
2.4	.30954	.30284	.28546	.23706	.18255	.14116	.12632	.11184
2.6	.27387	.26024	.24117	.19065	.14288	—	—	—
2.8	.23387	.21939	.19834	.14980	—	—	—	—
3.0	.19953	.18053	.15882	—	—	—	—	—
3.2	.16449	.14500	—	—	—	—	—	—
3.4	.13256	—	—	—	—	—	—	—

TABLE III

Values of the density of $z = x_{(0)} - cx_{(1)} - (1 - c)\,x_{(2)}$ *under the hypothesis* $H_2 : \mu_0 = 2\delta,\ \mu_1 = \delta,\ \mu_2 = 0;\ \delta = 1$

c	0	0.1	0.2	0.4	0.6	0.8	0.9	1.0
0.0	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	.54317
0.2	.04056	.03960	.05069	.06759	.10137	.20168	.38016	.52172
0.4	.08109	.09037	.10133	.13497	.20146	.37844	.51800	.49788
0.6	.12140	.13481	.15152	.20107	.29509	.47802	.50288	.47151
0.8	.16106	.17860	.20039	.26360	.37311	.49628	.47401	.44252
1.0	.19934	.22058	.24647	.31926	.42638	.47505	.44270	.41092
1.2	.23518	.25927	.28808	.36377	.45112	.44210	.40851	.37691
1.4	.26731	.29305	.32298	.39430	.44918	.40526	.37182	.34096
1.6	.29437	.32027	.34906	.40863	.42718	.36591	.33331	.30372
1.8	.31476	.33859	.36482	.40664	.39249	.32495	.29381	.26591
2.0	.32837	.34951	.36940	.38986	.35119	.28344	.25435	.22887
2.2	.33363	.35004	.36285	.36159	.30690	.23920	.21601	.19316
2.4	.33070	.34123	.34610	.32502	.26260	.19934	.17977	.15978
2.6	.31996	.32398	.32083	.28390	.21793	—	—	—
2.8	.30177	.29972	.28927	.24141	—	—	—	—
3.0	.27902	.27019	.25383	—	—	—	—	—
3.2	.25165	.23777	—	—	—	—	—	—
3.4	.22185	—	—	—	—	—	—	—

REFERENCES

[1] A. T. McKay, "The Distribution of the difference between the extreme observation and the sample mean in samples of n from a normal universe," *Biometrika*, Vol. 27 (1935), pp. 466–471.

[2] A. T. McKay and E. S. Pearson, "A note on the distribution of range in samples of size n," *Biometrika*, Vol. 25 (1933), pp. 415–420.

[3] National Bureau of Standards, "Tables of the exponential function e^x," Applied Math. Series, No. 14, 1951.

[4] National Bureau of Standards, "Tables of normal probability functions," Applied Math. Series, No 23, 1953.

[5] K. Pearson, *Tables for Statisticians and Biometricians*, Cambridge University Press, Part I, 1930.

[6] J. St-Pierre and A. Zinger, "The null distribution of the difference between the two largest sample values," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 849–851.

[7] S. S. Wilks, "Order statistics," *Bull. Amer. Math. Soc.*, Vol. 54 (1948), pp. 6–50.

ADMISSIBLE ONE-SIDED TESTS FOR THE MEAN OF A RECTANGULAR DISTRIBUTION[1]

By J. W. Pratt

Harvard University

1: Theorem. *Suppose we have a sample of $n > 1$ independent observations from a uniform distribution with unknown mean θ and known range R. Suppose we wish to test $H_0:\theta \leqq \theta_0$ against $H_1:\theta > \theta_0$. Then an essentially complete class of admissible tests is the class $\mathfrak{A}$ of all tests of the following type. Let u be the minimum observation, v the maximum. Let $g(u)$ be a nonincreasing function of u such that $g(u) = \theta_0 + \frac{1}{2}R$ for $u < \theta_0 - \frac{1}{2}R$. Accept H_0 if and only if $v < g(u)$.*

2. Discussion. The two-sided problem has been treated by Allan Birnbaum [1]. He showed that, for testing $H_0':\theta = \theta_0$ against $H_1':\theta \neq \theta_0$, an essentially complete class of admissible tests is the class of all tests of the following type. Let $v(u)$ be a nondecreasing function of u. Accept H_0 if and only if $v > v(u)$ and $\theta_0 - \frac{1}{2}R < u < v < \theta_0 + \frac{1}{2}R$.

Birnbaum [1] also noted that there is a uniformly most powerful size α test of $H_0':\theta = \theta_0$ against $H_1:\theta > \theta_0$, namely that accepting H_0' if $\theta_0 - \frac{1}{2}R < u < \theta_0 + (\frac{1}{2} - \alpha^{1/n})R$ and $v < \theta_0 + \frac{1}{2}R$. This corresponds in our notation to

$$g(u) = \begin{cases} \theta_0 + \frac{1}{2}R \text{ for } u < \theta_0 + (\frac{1}{2} - \alpha^{1/n})R, \\ \theta_0 - \frac{1}{2}R \text{ (say) otherwise.} \end{cases}$$

Received July 15, 1957; revised March 21, 1958.

[1] Research carried out at Stanford University and revisions at University of Chicago, and at Harvard University, under partial sponsorship of the Statistics Branch, Office of Naval Research. Reproduction in whole or in part is permitted for any purpose of the United States Government.

In this rather simple situation, then, an essentially complete class of admissible tests of the simple hypothesis against one-sided alternatives consists of the uniformly most powerful test (just described) for each significance level, but the class of admissible tests of the composite hypothesis against one-sided alternatives is very general. The class of admissible tests of the simple hypothesis against two-sided alternatives is also very general, but quite different. It includes unions of admissible lower and upper one-sided rejection regions (if and) only if they are admissible for the simple hypothesis, and such unions form a portion "of measure zero" in the whole class.

In the following section we will prove the result stated in the first paragraph. The proof uses no general results of decision theory, such as the complete class theorem, but only direct methods of an essentially elementary constructive type. It obviously works in some slightly more general situations, which are given explicitly in [2].

3. Proof. Without loss of generality we may take $\theta_0 = 0$, $R = 2$. Since (u, v) is a sufficient statistic, an essentially complete class of tests is the class of all randomized tests based on (u, v). Suppose such a test is given, accepting H_1 with probability $\phi_0(u, v)$ when (u, v) is observed.

The triangle $T(\theta) = \{(u, v): \theta - 1 < u \leqq v < \theta + 1\}$ contains (u, v) with probability one if θ is the true mean. The probability of accepting H_0 using the test function ϕ is

$$E_\theta(\phi) = \iint_{T(\theta)} \phi(u, v) 2^{-n} n(n-1)(v-u)^{n-2}\, du\, dv. \tag{1}$$

If $\theta \geqq 0$, then $u > -1$ with probability one. If $\theta \leqq 0$, then $v < 1$ with probability one. Thus if (u, v) is not in $T(0)$, we know which hypothesis is correct. Accordingly, let

$$\phi_1(u, v) = \begin{cases} \phi_0(u, v) & \text{if} \quad (u, v) \,\varepsilon\, T(0), \\ 1 & \text{if} \quad u \leqq -1, \\ 0 & \text{if} \quad v \geqq 1. \end{cases} \tag{2}$$

Then ϕ_1 dominates ϕ_0, i.e. ϕ_1 is at least as good as ϕ_0 for any θ, i.e.

$$E_\theta(\phi_1) \gtreqless E_\theta(\phi_0) \quad \text{for} \quad \theta \lesseqgtr 0. \tag{3}$$

Define $f(v)$ for $-1 < v < 1$ by

$$\int_{-1}^{f(v)} (v-u)^{n-2}\, du = \int_{-1}^{v} \phi_1(u, v)(v-u)^{n-2}\, du, \quad -1 \leqq f(v) \leqq v. \tag{4}$$

Let

$$\phi_2(u, v) = \begin{cases} 1 \text{ if } u \leqq f(v),\ -1 < v < 1, \text{ or if } v \leqq -1, \\ 0 \text{ if } u > f(v),\ -1 < v < 1, \text{ or if } v \geqq 1. \end{cases} \tag{5}$$

Then, with respect to the density $2^{-n}n(n-1)(v-u)^{n-2}du$, $v-1 < u < v$, ϕ_2 has the same mass as ϕ_1 on each horizontal line in the (u, v)-plane, but concentrates it as far to the left as possible. Furthermore, $\phi_2 = \phi_1$ except on $T(0)$. Therefore

$$E_\theta(\phi_2) \overset{=}{\leqq} E_\theta(\phi_1) \quad \text{for} \quad \theta \overset{\leqq}{>} 0. \tag{6}$$

Therefore ϕ_2 dominates ϕ_1.

Define $g(u)$ for $-1 < u < 1$ by

$$\int_u^{g(u)} (v-u)^{n-2}\, dv = \int_u^1 \phi_2(u, v)(v-u)^{n-2}\, dv, \quad u < g(u) \leqq 1, \tag{7}$$

if the right-hand side is positive. If the right-hand side vanishes, or if $u \geqq 1$, let $g(u) = -1$. If $u \leqq -1$, let $g(u) = 1$. Let

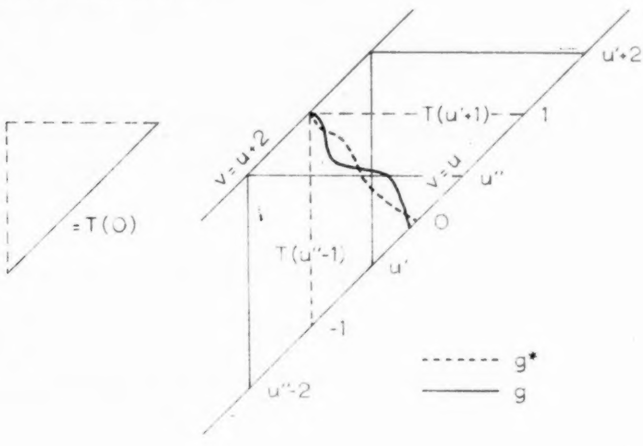
$$\phi_3(u, v) = \begin{cases} 1 & \text{if} \quad v < g(u), \\ 0 & \text{if} \quad v \geqq g(u). \end{cases} \tag{8}$$

Then, with respect to the density $2^{-n}n(n-1)(v-u)^{n-2}\,dv$, $u < v < u+1$, ϕ_3 has the same mass as ϕ_2 on each vertical line in the (u, v)-plane, but concentrates it as low as possible. Furthermore, $\phi_3 = \phi_2$ except on $T(0)$. Therefore

$$E_\theta(\phi_3) \overset{\geq}{=} E_\theta(\phi_2) \quad \text{for} \quad \theta \overset{<}{\geqq} 0. \tag{9}$$

Therefore ϕ_3 dominates ϕ_2.

By (5), $\phi_2(u, v)$ is nonincreasing in u for each v. Therefore, by (7), for $-1 < u < g(u)$, $-1 < g(u) \leqq 1$, and $g(u)$ is nonincreasing in u. This is the essential part of the requirement that ϕ_3 be in $\mathfrak{A}$, and $g(u)$ was defined for other values of u so that ϕ_3 actually is in $\mathfrak{A}$.



We have thus shown that any test is dominated by a test in $\mathfrak{A}$, i.e. that $\mathfrak{A}$ is essentially complete. It remains to prove admissibility. Suppose ϕ and ϕ^* are given by g and g^*. Without changing the characteristics of the tests, we may redefine g and g^* so that they are left-continuous and so that $g(u) = -1$ where $g(u) \leqq u$, and $g^*(u) = -1$ where $g^*(u) \leqq u$. Suppose there is a u' such that $g(u') > g^*(u')$. Choose u'' such that $g(u') > u'' > g^*(u')$. (See the diagram.) Let "area" be measured with respect to the density $2^{-n}n(n-1)(v-u)^{n-2}du\,dv$. By left-continuity, $g^*(u) < u$ for all u in an interval whose right endpoint is u'. Therefore either the "area" below g in $T(u' + 1)$ is less than that below g^*, or the "area" below g in $T(u'' - 1)$ is greater than that below g^*. But the "area" below g in $T(\theta)$ is just $E_\theta(\phi)$. Thus either $E_{u'+1}(\phi) < E_{u'+1}(\phi^*)$ or $E_{u''-1}(\phi) > E_{u''-1}(\phi^*)$. But $u' + 1 > 0$ and $u'' - 1 < 0$, so this shows ϕ doesn't dominate ϕ^*. Hence if ϕ dominates ϕ^*, $g(u') \leqq g^*(u')$ for all u'. But in this case either ϕ and ϕ^* are essentially the same or $E_\theta(\phi) < E_\theta(\phi^*)$ for sufficiently small positive θ. Therefore ϕ cannot dominate ϕ^*. Since ϕ and ϕ^* were arbitrary tests of the essentially complete class $\mathfrak{A}$, it follows that all tests in A are admissible.

This proof of admissibility is spelled out analytically in [2]. The proof of essential completeness given there uses a general property possessed by the rectangular distribution.

REFERENCES

[1] A. Birnbaum, "Admissible tests for the mean of a rectangular distribution," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 157-161.

[2] J. W. Pratt, "Some results in the decision theory of one-parameter multivariate Polya type distributions," Stanford Technical Report No. 37 (Oct. 21, 1955), pp. 45-50.

A METHOD FOR SELECTING THE SIZE OF THE INITIAL SAMPLE IN STEIN'S TWO SAMPLE PROCEDURE

By Jack Moshman

Corporation for Economic and Industrial Research, Arlington 2, Virginia

1. Summary and Introduction. The use of an upper percentage point of the distribution of total sample size in conjunction with the expectation of the latter is proposed as a guide to the selection of the size of the initial sample when using some version of Stein's [5] two-sample procedure. It is a rapidly calculable function of the underlying population variance based on existing tables of the χ^2 distribution. A rule-of-thumb is proposed to be used in making the actual selection of initial sample size. It is a simple matter to investigate the nature of the percentage point for different values of the variance over a limited range;

Received July 22, 1957; revised December 2, 1957.

a recommended conservative choice when the variance is not known is the selection of a large initial sample.

Dantzig [2] proved the nonexistence of nontrivial tests of Student's hypothesis whose power was independent of the variance, a result extended by Stein to the general linear hypothesis. In the same paper Stein proposed a two-sample procedure whose power was independent of variance. The same two-sample method could be used to obtain a confidence interval for the mean of a normal distribution with predetermined length and confidence coefficient.

Stein gave no specifications for the choice of the initial sample size, but Seelbinder [4] suggested that it be selected to minimize the expectation of the total sample. In a recent paper, Bechhofer, Dunnett and Sobel [1] used Stein's procedure for another application, noting that the variance of the total sample size increased as the size of the first sample decreased.

An efficient choice of the size of the initial sample will hold the expectation of the sample small, and will further reduce the probability of an extremely large total sample. This note will explore the matter in further detail and show that an upper percentage point of the distribution of total sample size, when used in conjunction with the expectation, is a rapidly calculable guide to an efficient choice of the size of the first sample.

2. Basic theory. As developed by Stein, the two-sample procedure involves a preliminary, arbitrary choice of a positive integer N_0 and a number $z > 0$. The value of z will depend, when constructing a confidence interval of length $2L$ for the mean, on the precision of the estimate, i.e., the length of the interval, and its reliability, the confidence coefficient. Specifically, if $t_{n,\gamma}$ is the upper 100 γ percentage point of Student's distribution with n degrees of freedom, one would take $z = L^2/t^2_{N_0-1,1-(\alpha/2)}$ to obtain a confidence coefficient $\geqq 1 - \alpha$.

A sample of N_0 observations is taken and $s^2 = \sum(x_i - \bar{x})^2/(N_0 - 1)$ is computed as an estimate of the unknown variance σ^2 with $n = N_0 - 1$ degrees of freedom. The total sample size, N, is then

$$N = \max\left(\left[\frac{s^2}{z}\right] + 1, N_0\right), \tag{1}$$

where $[t]$ is the *largest integer less than t.*

Hence it follows that

$$\text{Prob}\,(N = N_0) = \text{Prob}\left(\frac{s^2}{z} \leqq N_0\right) = \text{Prob}\left(\frac{ns^2}{\sigma^2} = \chi^2(n) \leqq \frac{nN_0 z}{\sigma^2}\right), \tag{2}$$

where $\chi^2(n)$ is distributed as χ^2 with n degrees of freedom. Furthermore, for integral $m > N_0$,

$$\begin{aligned} \text{Prob}\,(N = m) &= \text{Prob}\left(m - 1 < \frac{s^2}{z} \leqq m\right) \\ &= \text{Prob}\left(\frac{n(m-1)z}{\sigma^2} < \chi^2(n) \leqq \frac{nmz}{\sigma^2}\right). \end{aligned} \tag{3}$$

Therefore, letting $\lambda = z/\sigma^2$, one may easily show

$$(4)\quad \begin{aligned} E(N) = N_0 \operatorname{Prob}(\chi^2(n) < n\lambda N_0) + \frac{1}{\lambda} \operatorname{Prob}(\chi^2(n+2) > n\lambda N_0) \\ + \theta_1 \operatorname{Prob}(\chi^2(n) > n\lambda N_0) \end{aligned}$$

and

$$(5)\quad \begin{aligned} \operatorname{Var}(N) = N_0^2 \operatorname{Prob}(\chi^2(n) < n\lambda N_0) + \frac{(n+2)}{n\lambda^2} \operatorname{Prob}(\chi^2(n+4) > n\lambda N_0) \\ + \frac{2\theta_2}{\lambda} \operatorname{Prob}(\chi^2(n+2) > n\lambda N_0) + \theta_3 \operatorname{Prob}(\chi^2(n) > n\lambda N_0) - (E(N))^2, \end{aligned}$$

where $0 \leqq \theta_i \leqq 1$, $i = 1, 2, 3$.

Whereas (4) defines $E(N)$ within a maximum error of unity, (5) is not as useful inasmuch as the factor $1/\lambda$ may be, and frequently is, large.

Furthermore, it is somewhat difficult to translate Var (N) into working percentage points of the distribution of N. A more useful procedure is to calculate a given percentage point N_p of the distribution. This may be accomplished directly from (2) and (3). Define N_p as the smallest integer $\geqq N_0$ such that

$$(6)\quad \operatorname{Prob}(N \leqq N_p) = \sum_{m=N_0}^{N_p} \operatorname{Prob}(N = m) \geqq p.$$

But this is equivalent, if one writes $p_n(\chi^2)$ as the probability density function of $\chi^2(n)$, to setting

$$(7)\quad \int_0^{nN_p\lambda} p_n(\chi^2)\, d\chi^2 \geqq p$$

and letting N_p be chosen to satisfy (7), but not less than N_0. Thus

$$(8)\quad N_p = \max\left\{N_0, \left[\frac{1}{\lambda}\left(100p\text{th } \textit{percentage point of } \frac{\chi^2(n)}{n}\right)\right] + 1\right\},$$

which is tabulated in Hald [3] for example. Note that the upper percentage points of $\chi^2(n)/n$ decreases monotonically as n increases. Conceivably, if N_0 is chosen very large, one can be reasonably confident that no further sampling will be necessary, but this is not an efficient procedure.

A rough, but objective, rule-of-thumb may be derived by the following consideration: Let $E(N \mid N_0^*)$ be the expectation of N if $N_0 = N_0^*$ and $N_p(N_0^*)$ the 100pth percentile of N if $N_0 = N_0^*$. Define

$$(9)\quad P(N_0^*) = \int_0^{n\lambda E(N\mid N_0^*)} p_n(\chi^2)\, d\chi^2,$$

as the proportion of time N will not exceed $E(N \mid N_0^*)$. Let $\mathbf{N}_0$ be the value of N_0 which minimizes $E(N)$, i.e.,

$$(10)\quad E(N \mid \mathbf{N}_0) \leqq E(N \mid N_0)$$

for all N_0. Now one might investigate alternative values of N_0 by considering

$$\Psi(N_0) = (1 - p)(N_p(\mathbf{N}_0) - N_p(N_0)) - (1 - P(\mathbf{N}_0))(E(N \mid N_0) - E(N \mid \mathbf{N}_0)) \tag{11}$$

and selecting N_0 as the integer for which $\Psi(N_0)$ is a maximum. In effect, (11) weights the expected changes in $E(N)$ and N_p by the probability of exceeding those values. It would be expected that p would be chosen independently from nonstatistical considerations.

3. Example. If one takes $\lambda = .1$, where in the ordinary application considered by Stein $n = N_0 - 1$, then $E(N)$ is a minimum for $N_0 = \mathbf{N}_0 = 3$. Values of $E(N \mid N_0)$ are tabulated in Table 1. It may be seen that $E(N \mid N_0)$ is fairly constant over a considerable range. The same table also contains $N_{.95}(N_0)$ which decreases sharply where $E(N \mid N_0)$ is relatively constant.

It may readily be verified from (9) that $P(N_0) = P(3) \approx .64$. Rapidly one may evaluate $\Psi(N_0)$ from (11), taking $p = .95$, and find that $\Psi(6) = .2686$ is the maximum. Hence the rule suggested specifies $N_0 = 6$ as the proper choice.

4. Discussion. When the variance is unknown, two alternatives exist. It may be feasible to express the length of the confidence interval desired as a proportion of σ; no difficulty then ensues since λ is specified. If L is specified absolutely, in most practical cases a range for σ is known. One can then investigate the distribution of N for various values of σ in this range and make a subsequent choice of N_0.

The procedures suggested in this note are particularly applicable to those situations where repeated sampling is not contemplated and/or there exists a physical reason for wanting to avoid excessively large samples. The latter situation may obtain where larger individual samples may entail the purchase of additional test equipment or require the supplementing of a regular interviewing staff by extra employees.

TABLE 1
Dependence of $E(N)$ and $N_{.95}$ on N_0
$\lambda = .1$

N_0	$E(N \mid N_0)$	$N_{.95}(N_0)$	N_0	$E(N \mid N_0)$	$N_{.95}(N_0)$
2	10.45	38.41	10	11.84	18.80
3	10.29	29.96	12	12.92	17.89
4	10.45	26.05	14	14.35	17.20
5	10.51	23.72	16	16.15	16.66
6	10.63	22.14	18	18.02	18.00
7	10.80	20.99	20	20.02	20.00
8	11.18	20.10	22	22.01	22.00
9	11.43	19.38	24	24.00	24.00

5. Acknowledgment. The author wishes to express his indebtedness to Professor G. E. Albert for many helpful suggestions made in the pursuance of this research.

REFERENCES

[1] R. E. Bechhofer, C. W. Dunnett and M. Sobel, "A two-sample multiple decision procedure for ranking means of normal populations with a common unknown variance," *Biometrika*, Vol 41 (1954), p. 170.

[2] G. B. Dantzig, "On the non-existence of tests of 'Student's' hypothesis having power functions independent of σ," *Ann. Math. Stat.*, Vol 11 (1940), p. 186.

[3] Anders Hald, *Statistical Theory with Engineering Applications*, John Wiley & Sons, 1952.

[4] B. M. Seelbinder, "On Stein's two-stage sampling scheme," *Ann. Math. Stat.*, Vol. 24 (1953), p. 640.

[5] Charles Stein, "A two-sample test for a linear hypothesis whose power is independent of the variance," *Ann. Math. Stat.*, Vol. 16 (1945), p. 243.

ON A PROBLEM IN MEASURE-SPACES

By V. S. Varadarajan

Indian Statistical Institute, Calcutta

Summary. Let $\mathfrak{F}$ be the family of all random variables on a probability space Ω taking values from a separable and complete metric space X. In this paper we prove that $\mathfrak{F}$ is in a certain sense a closed family. More precisely, if $\{\xi_n\}$ is a sequence of X-valued random variables such that their probability distributions converge weakly to a probability distribution P on X, then there exists an X-valued random variable on Ω with distribution P. An example is also given which shows that the assumption of completeness of X cannot in general be dropped.

1. Preliminary remarks. In what follows $(\Omega, \mathcal{S}, \mu)$ is a probability space and X a separable metric space. We denote by $\mathcal{B}$ the class of Borel subsets of X defined as the minimal σ-field containing all open subsets of X.

A map φ of Ω into X is called a random variable if it is measurable i.e., $\varphi^{-1}(A) \varepsilon \mathcal{S}$ for each $A \varepsilon \mathcal{B}$. If φ is a random variable we define as its distribution the measure μ_φ on $\mathcal{B}$ given by

$$\mu_\varphi(A) = \mu\{\varphi^{-1}(A)\}$$

for all $A \varepsilon \mathcal{B}$. A given probability measure P on $\mathcal{B}$ is said to be induced from Ω if there exists a random variable φ such that $P = \mu_\varphi$.

Suppose we are given a sequence $\{P_n\}$ of probability measures on $\mathcal{B}$. We say that $\{P_n\}$ converges weakly to a probability measure P on $\mathcal{B}$ ($P_n \Rightarrow P$ in symbols) if

$$\lim_{n\to\infty} \int_X g\, dP_n = \int_X g\, dP$$

Received November 12, 1956; revised May 19, 1958.

for every bounded continuous function g on X. In terms of subsets of X this is equivalent to

$$\limsup_{n\to\infty} P_n(C) \leqq P(C)$$

for every closed set $C \subset X$ ([1]). When X is the real line with the usual topology, this convergence is equivalent to the usual convergence of distributions.

2. The main theorem. In this section we state and prove the main theorem. Before doing it we prove a lemma.

LEMMA. *Let X be a separable and complete metric space and $(\Omega, \mathcal{S}, \mu)$ a nonatomic probability space* ([2] p. 168). *Then any probability measure on $\mathcal{B}$ can be induced from Ω.*

PROOF. Since X is a separable metric space, it can be imbedded homeomorphically into a countable product of unit intervals by a celebrated theorem of Urysohn ([3] p. 125). Since it is also complete, the image of X will be a G_δ by a theorem of Larentieff ([3] p. 207). X can thus be regarded as a Borel subset of a countable product of unit intervals. This implies however that X can be regarded as a Borel subset of the unit interval since the unit interval and the countable product of such intervals can be connected by an one-one map which is measurable both ways. It is thus sufficient to show that any probability measure on the unit interval can be induced from Ω. This however is a well-known result.

We now prove the main theorem.

THEOREM. Let X be a separable and complete metric space and $(\Omega, \mathcal{S}, \mu)$ an arbitrary probability space. If $\{\xi_n\}$ is a sequence of X-valued random variables such that $\mu_{\xi_n} \Rightarrow P$ as $n \to \infty$ where P is a probability measure on $\mathcal{B}$, there exists an X-valued random variable ξ such that $P = \mu_\xi$.

PROOF. Any measure space can be decomposed into its atomic and nonatomic components and in view of the previous lemma we can assume that there is no nonatomic component in Ω. We can thus write $\Omega = A_1 \cup A_2 \cup \cdots$ where (i) $A_i \cap A_j = \phi$ for $i \neq j$, (ii) each A_i is an atom of $(\Omega, \mathcal{S}, \mu)$, and (iii) $\mu(A_i) = c_i > 0$ for each i. The distribution $P_n(\equiv \mu_{\xi_n})$ is then atomic and (since X is separable) has mass concentrated in a countable set of points, say $\{a_{n1}, a_{n2}, \cdots\}$. $P_n[a_{ni}] = c_i$ for $i = 1, 2, \cdots$.

We first assert that for each i, the set $D_i = \{a_{1i}, a_{2i}, \cdots\}$ has compact closure. If not, then for some i_0, D_{i_0} has a subset which has no limit point and which is infinite. We can assume without losing generality that this subset is D_{i_0} itself and that all the a_{ni_0} are distinct. If then $D \subset D_{i_0}$ is any subset, then D is closed and from $P_n \Rightarrow P$ it follows that $P(D) \geqq \limsup_{n\to\infty} P_n(D)$. If D is infinite then, $\limsup_{n\to\infty} P_n(D) \geqq c_{i_0}$. Thus for any infinite subset $D \subset D_{i_0}$, $P(D) \geqq c_{i_0} > 0$ which is a contradiction.

Thus each D_i has compact closure. We can then, by the diagonal procedure choose a sequence $\{n_k\}$ of integers and points $a_1, a_2, \cdots$ of X such that

$$\lim_{k\to\infty} a_{n_k, i} = a_i$$

for $i = 1, 2, \cdots$. Let ξ be the random variable with values $a_1, a_2, \cdots$ on the sets $A_1, A_2, \cdots$. We complete the proof by showing that $P = \mu_\xi$. It is enough to show that $P_{n_k} \Rightarrow \mu_\xi$. In fact for any bounded continuous g on X,

$$\int_X g\, dP_{n_k} = \sum_i c_i\, g(a_{n_k, i}) \to \sum_i c_i\, g(a_i) = \int_X g\, d\mu_\xi,$$

the passage to the limit being justified as $\sum_i c_i g(a_{n_k,i})$ converges uniformly in k. This completes the proof of the theorem.

Remarks. (1) Suppose X is any separable metric space and X^* its completion. The above theorem will still be true not for X but for X^* and ξ will now be X^*-valued. If then X has the property that as a subset of X^* it is measurable with respect to the completion of *every* measure on X^*, ξ can be reduced to an X-valued random variable and the main theorem is true for such X. This is the case for instance when X is itself a Borel set in X^*. It is interesting to note that there are separable metric spaces X which have the above mentioned property in relation to X^* but which are not complete under any metrization, for example, the set of rationals with the relative real line topology.

(2) It is to be noted that when $(\Omega, \mathcal{S}, \mu)$ is purely atomic the theorem is true with any separable X.

(3) Suppose now $A_1, A_2, \cdots$ is a sequence of sets in $\mathcal{S}$ such that $\mu(A_n) \to \alpha$. Setting $\xi_n = \chi_{A_n}$, the characteristic function of A_n, we find that $\mu_{\xi_n} \Rightarrow P$ where P is the measure with masses α and $1 - \alpha$ at the points 1 and 0. The above theorem then ensures the existence of $A \varepsilon \mathcal{S}$ such that $\mu(A) = \alpha$; in other words that the range of μ is a closed subset of $[0, 1]$.

3. An example. We construct an example to show that the theorem proved in Section 2 requires some such condition on X. We take for X a subset of $[0, 1]$ such that (i) $\mu^*(X) = 1$, $\mu_*(X) = 0$ where μ is Lebesgue measure and (ii) X contains all points of the form $m/2^n$. For $(\Omega, \mathcal{S}, \mu)$ we take the unit interval with Lebesgue measure. The Borel sets of X are precisely the intersections with X of Borel subsets of $[0, 1]$. Lebesgue outer measure on $\mathcal{B}$ is now actually a measure over it, denoted by λ.

Suppose now P_n is the measure on $\mathcal{B}$ with equal masses $1/2^n$ at the points $m/2^n$ $(m = 1, 2, \cdots 2^n)$. It is easy to verify that $P_n \Rightarrow \lambda$. Further each P_n is trivially induced from Ω. We will now show that λ cannot be induced from Ω.

Suppose λ is induced by the map ξ. ξ is obviously a Borel measurable function on $[0, 1]$ and hence by Lusin's theorem ([2]) p.243) we can find for each $\epsilon > 0$ a compact $K_\epsilon \subset [0, 1]$ such that (i) $\mu(K_\epsilon) > 1 - \epsilon$ and (ii) ξ restricted to K_ϵ is continuous. If $M_\epsilon = \xi[K_\epsilon]$, then $M_\epsilon \subset X$ and is a compact subset of the real line. Since λ is induced by ξ, $\lambda(M_\epsilon) > 1 - \epsilon$. But M_ϵ is a Borel set of the real line and this shows that $\mu(M_\epsilon) > 1 - \epsilon$, contradicting the assumption that $\mu_*(X) = 0$.

Thus λ cannot be induced from Ω. This completes the discussion of the example.

REFERENCES

[1] P. Billingsley, "The invariance principle for dependent random variables," *Trans. Amer. Math. Soc.* (1956).
[2] P. R. Halmos, *Measure Theory*, D. Van Nostrand, New York, 1950.
[3] J. L. Kelley, *General Topology*, D. Van Nostrand, New York, 1955.

CORRECTION TO "PROBABILITIES OF HYPOTHESES AND INFORMATION-STATISTICS IN SAMPLING FROM EXPONENTIAL-CLASS POPULATIONS"

By Morton Kupperman

The George Washington University

In the paper cited in the title (*Ann. Math. Stat.*, Vol. 29 (1958), pp. 571–575): p. 572, line 5. For $\sum \mathbf{x}p(\mathbf{x}, \boldsymbol{\theta}_m)$ read $\sum_{\mathbf{x}} p(\mathbf{x}, \boldsymbol{\theta}_m)$.

CORRECTION TO "POWER FUNCTIONS OF THE GAMMA DISTRIBUTION"

G. D. Berndt

Professor I. R. Savage has called to my attention, through the Editor, the fact that I have overlooked reference to previous work appearing in Eisenhart, Haystay, and Wallis, *Techniques of Statistical Analysis*, and bearing on results reported by me in the *Annals*, Vol. 29, No. 1, March 1958, pages 302–306.

On pages 274–275 of Eisenhart, Haystay, and Wallis, in Figures 8.1 and 8.2, there are given operating characteristic curves for the chi-squared distribution for eight selected degrees of freedom when the significance level is 0.01 and 0.05. Inasmuch as the chi-squared distribution is a gamma distribution with $\frac{1}{2}$ (degrees of freedom) = the parameter gamma in my paper and with 2 = the parameter beta in my paper, and since their rho is equivalent to my delta, there is a similarity in the reported results. This similarity has resulted in some overlap in the results of the two papers in that ten of my forty-eight power curves have an equivalent in the operating characteristic curves in the previous work.

I should like to acknowledge this previous work, and also that of Ferris, Grubbs, and Weaver, by having the following two references added to the two which already appear at the end of my paper:

[3] *Selected Techniques of Statistical Analysis*, Churchill Eisenhart, Millard W. Haystay, and W. Allen Wallis, editors, McGraw-Hill, New York, 1947, pp. 270–278.
[4] Charles D. Ferris, Frank E. Grubbs, and Chalmers L. Weaver, "Operating characteristics for some common statistical tests of significance," *Annals of Mathematical Statistics*, Vol. 17 (1946), pp. 178–197.

ABSTRACTS OF PAPERS

(Abstracts of papers presented for the Cambridge, Massachusetts Meeting of the Institute, August 25-28, 1958.)

48. **On the Relationship Algebra and the Association Algebra of the Partially Balanced Incomplete Block Design.** JUNJIRO OGAWA, University of North Carolina.

A. T. James (1957) defined the so-called "relationship algebra of a design" and showed that the partition of the total sum of squares into partial sums of squares can be characterized by the structure of the relationship algebra. He constructed the relationship algebras and analyzed their algebraic structures for the randomized block, the latin-square and the balanced incomplete block design. The purpose of this note is to construct the relationship algebra for the partially balanced incomplete block design and analize its algebraic structure. The main result of this paper is that the second degree irreducible representations of the relationship algebra are completely determined by the irreducible (linear) representations of the association algebra defined by R. C. Bose. (Received June 2, 1958.)

49. **Estimation of the Medians for Dependent Variables.** OLIVE JEAN DUNN, University of California.

The problem considered in this paper is that of using non-parametric methods to estimate the unknown medians of two dependent variables. In various types of research, it is convenient to consider a sample of n individuals and to take measurements at two different times or at two different levels of treatment. These $2n$ measurements are then a sample of size n from a bivariate distribution. For independent variables, two confidence intervals of the classic type using order statistics may be used as simultaneous confidence intervals for the two medians by simply multiplying the two probabilities to obtain the new confidence level. In this paper it is shown that for two dependent variables these same confidence intervals may be used as a set with bounded confidence level. Comparisons are made on the basis of average length between these intervals and other joint intervals for the means of a bivariate normal distribution. It is also shown that the result of this paper does not generalize, at any rate in the most obvious way, to three or more dependent variables. (Received June 16, 1958.)

50. **On the Problem of Incomplete Data.** (Preliminary report) JUNJIRO OGAWA and BERNARD S. PASTERNACK, University of North Carolina.

Consider a sample of size n, $x_1, x_2, \cdots, x_n$, drawn from $N(\mu, \sigma^2)$ (σ^2 known). Suppose only $n - k$ observations are available, $x_1, x_2, \cdots, x_{n-k}$ (say). Let

$$\bar{x} = 1/n \sum_{i=1}^{n} x_i,\ \bar{x}^* = 1/(n-k) \sum_{i=1}^{n-k} x_i,\ u = (\sqrt{n}\,\bar{x})/\sigma,\ u^* = (\sqrt{n-k}\,\bar{x}^*)/\sigma$$

and define $u(\alpha)$ by $P_{H_0}(|u| \leq u(\alpha)) = 1 - \alpha$ and $u^*(\alpha)$ by $P_{H_0}(|u^*| \leq u^*(\alpha)) = 1 - \alpha$, where $H_0{:}\mu = \mu_0$. We define $P_H\{|u| \leq u(\alpha) \mid |u^*| > u^*(\alpha)\}$, α being prefixed, as the *reversal function* of this test procedure. The reversal function has been tabulated for various values of k/n. When σ^2 is unknown, the test procedure depends upon t, or, for incomplete data, t^*. A least upper bound for t given t^* has been obtained, i.e., the minimum value of τ such that $P_{H_0}(|t| > \tau \mid t^*) = 0$. Similar bounds (both l.u.b. and g.l.b.), in probability, have also been obtained for situations involving one-way classifications, the general linear model and Hotelling's T^2. Another approach to the problem of missing data involves the

introduction of a chance mechanism according to which observations are missed. Research along these lines is now in progress and the authors hope to present some of these results in the near future. (Received June 20, 1958.)

51. Aids for Fitting the Pea son Type III Curve by Maximum Likelihood. (Preliminary report) J. ARTHUR GREENWOOD, Iowa State College and DAVID DURAND, M.I.T.

New tables and formulas of approximation are given for the function $\rho = y\phi(y)$, where $\phi(y)$ is the inverse function to $y = \ln\rho - d/d\rho \ln\Gamma(\rho)$. With the aid of these tables, one may obtain by direct interpolation the maximum likelihood estimate (joint) of the exponent in a Type III distribution with known lower limit. Application of the tables to the Type III with unknown lower limit and to the Type V are briefly discussed. (Received June 20, 1958.)

52. Admissible Estimates and Maximum Likelihood Estimates (Preliminary report) ALLAN BIRNBAUM, Columbia University.

A definition of admissibility of a point-estimate of a real-valued parameter θ is formulated on the basis of a slightly generalized form of the Neyman-Pearson theory of confidence regions, using *Ann. Math. Stat.*, Vol. 27 (1956), pp. 544–545, without introduction of loss functions. Necessary and sufficient conditions for existence of such estimates are given under mild regularity conditions. By extending methods developed in *Ann. Math. Stat.*, Vol. 26 (1955), pp. 21–36, it is shown that each admissible estimate is obtainable as the (unique) solution of an equation $\partial/\partial\theta \log L(x, \theta) = G(\theta)$, where $G(\theta)$ is a known function and $L(x, \theta)$ is the likelihood function. Setting $G(\theta) = 0$ gives the maximum likelihood estimate $\hat{\theta}$, which is thus shown to be admissible. In the case of non-existence of admissible estimates, asymptotically admissible estimates are defined and shown under certain conditions to exist and to include $\hat{\theta}$. An estimate $\tilde{\theta}$ is called median-unbiased if Prob $\{\tilde{\theta} \leqq \theta \mid \theta\} = \frac{1}{2}$ for all θ. $\hat{\theta}$ is shown under general conditions to be asymptotically median-unbiased, and to be a convenient approximation (often close for moderate sample sizes) to the median-unbiased admissible estimate (which is often difficult to compute). Relations to sufficiency and to multi-parameter estimation problems are discussed. (Received June 24, 1958.)

53. Stochastic Models for the Electron Multiplier Tube (Preliminary report) EDWARD K. DALTON, WILLARD D. JAMES AND HOWARD G. TUCKER, University of California.

Four stochastic models are proposed for the electron multiplier tube, two being branching processes involving Poisson distributions and two being branching processes involving binomial distributions. In each model there are two unknown parameters. It is desired to determine the best model among these and to estimate the parameters for it. Although the probability generating functions in each case are easy to derive, explicit formulas for the probability distributions in each case could not be found. A method for testing these models is presented which is based on the following theorem. THEOREM. *Let X be a random variable which takes on non-negative integer values, and let $X_1, \cdots, X_n, \cdots$ denote an infinite sequence of independent observations on X. Let $g(u \mid \alpha) = E(u^X)$ be the probability generating function of X which depends on a (vector) parameter α and is continuous in α. Let α_0 be the true value of α, and assume that there exists a sequence $\{\hat{\alpha}_n\}$ of random variables which converges to α_0 with probability one. Then for any value of u for which $u^2 - u \neq 0$ and $g(u^2 \mid \alpha) < \infty$ there exists a subsequence $\{\hat{\alpha}_{N_n}\}$ of $\{\hat{\alpha}_n\}$ such that the*

limiting distribution of the ratio of $\sum\{u^{X_k} \mid 1 \leq k \leq n\} - ng(u \mid \hat{a}_{N_n})$ *to either the square root of* $n\{g(u^2 \mid \hat{a}_{N_n}) - g^2(u \mid \hat{a}_{N_n})\}$ *or to the square root of*

$$n(n^{-1}\sum\{u^{2X_k} \mid 1 \leq k \leq n\} - (n^{-1}\sum\{u^{X_k} \mid 1 \leq k \leq n\})^2)$$

is normal with mean zero and variance one. A resumé of numerical results is included for three different sets of data corresponding to three different energy inputs. (Received June 26, 1958.)

54. On the Choice of Sample Size in the Kolmogorov-Smirnov Tests. JUDAH ROSENBLATT, Purdue University.

If F_n is the empirical distribution based on independent random variables $X_1, \cdots, X_n$, with common c.d.f. F, it is well known that a test of the hypothesis H_0: $F = F_0$ having asymptotic probability of type one error not exceeding α is to reject H_0 if and only if $n^{1/2} d_1(F_n, F_0) = n^{1/2} \sup_x | F_n(x) - F_0(x) | > h_{1\alpha}$, where

$$\lim_{n\to\infty} P_F\{n^{1/2} d_1(F_n, F) > h_{1\alpha}\} = \alpha$$

if F is continuous. Massey has suggested that the sample size n needed to achieve

$$P_F\{\text{Reject } H_0\} \geq 1 - \beta \text{ when } d_1(F_0, F) \geq l$$

be chosen as follows: n is the smallest integer such that $2[n^{1/2} l - h_{1\alpha}] \geq \varphi_\beta$, where

$$\int_{-\infty}^{\varphi_\beta} (1/2\pi)^{1/2} e^{-t^2/2}\, dt = 1 - \beta.$$

This suggestion is motivated by the normal approximation to the binomial distribution. A thorough investigation is made of this suggested procedure, and a completely justified, still rather simple technique is devised for choosing n such that

$$P_F\{\text{Reject } H_0\} \geq 1 - \beta \text{ when } d_1(F_0, F) \geq l.$$

The investigation is in two parts. First a region (near $p = \frac{1}{2}$) is determined where

$$\sum_{\nu=0}^{[n(p+l)-n^{1/2}h_{1\alpha}]} \binom{n}{\nu} p^\nu (1-p)^{n-\nu}$$

takes on its minimum value. This, together with the Uspensky version of the normal approximation to the binomial (with correction and error term) leads to the justified procedure for choosing n with the desired properties. This n is not much larger than that suggested by Massey and is far smaller than the one derivable from Chebychev's inequality. (Received July 2, 1958.)

55. The Use of Sample Quasi-Ranges in Estimating Population Standard Deviation. H. LEON HARTER, Wright Air Development Center.

The use of sample quasi-ranges in estimating the standard deviation of normal, rectangular, and exponential populations is discussed. For the normal population, the expected value, variance, and standard deviation of the rth quasi-range for samples of size n are tabulated for $r = 0(1)8$ and $n = (2r + 2)(1)100$. The efficiency of the unbiased estimate of population standard deviation based on one sample quasi-range is tabulated for the same values of r and n. Estimates based on a linear combination of two quasi-ranges are considered, and a method is given for determining the weighting factor which maximizes the efficiency. The most efficient unbiased estimates based on one quasi-range for $n = 2(1)100$ and on linear combinations of two adjacent quasi-ranges and of any two quasi-

ranges ($r < r' \leq 8$) for $n = 4(1)100$ are tabulated, along with their efficiencies. An example illustrates the use of these estimates. For rectangular and exponential populations, the most efficient unbiased estimates based on one quasi-range are tabulated, together with their efficiencies, also the bias when estimates which assume normality are used. (Received July 2, 1958.)

56. On a Limiting Distribution Due to Renyi. D. G. CHAPMAN, University of Washington.

Let X be a real valued random variable with distribution function (d.f.) $F(x)$. Let $F_n(x)$ denote the emperical d.f. based on n independent observations $x_1, x_2, \cdots, x_n$ of X. Renyi ("On the theory of order statistics," *Acta Math.*, Acad. Sci. Hungary, Vol. 4 (1953), pp. 191–231) has given the limiting distribution of $n^{1/2}R_n(a) = n^{1/2} \sup_{F(x) \geq a} [F_n(x) - F(x)]/F(x)$ as n tends to infinity, a being an arbitrary positive constant. It is therefore of interest to determine the limiting distribution of $R_n(0)$, i.e., without the arbitrary restriction $F(x) \geq a$. The result is obtained that $P_r[R_n(0) \leq \epsilon] = \epsilon/1 + \epsilon$ for all n, so that the limiting distribution of $R_n(0)$ has the same form. Also studied in this paper are the limiting distributions of some slight generalizations of $R_n(a)$. The method used is that due to Doob which is simpler than Renyi's and may also be used to determine the asymptotic power of the Smirnov test of goodness-of-fit for certain alternatives. (Received July 2, 1958.)

57. Power and Control of Size of Some Optimal Welch-type Statistics. ROGER S. MCCULLOUGH AND JOHN GURLAND.

A Welch-type statistic (Welch, *Biometrika*, 1938) is considered for testing equality of means in two normal populations with unknown variances which may be unequal. For various combinations of small sample sizes a nearly perfect one-sided control of size is possible, that is, optimal statistics are available which keep the size extremely close to a preassigned level if one population has a larger variance than the other. For two-sided control of size, that is with no restriction on the direction of inequality of variances, optimal statistics are available which keep the size below a pre-assigned level but arbitrarily close to the level over an infinite range of variance values. A table giving the optimal statistics for various combinations of small sample sizes has been prepared with the aid of an electronic computer. Tables of the power are also included. (Received July 2, 1958.)

58. A Note on Estimating Translation and Scalar Parameters. JOSEPH A. DUBAY, University of Oregon.

Let $X = (X_1, \cdots, X_n)$ be a random variable whose distribution depends on an unknown real valued parameter θ. Let $\delta(X)$ be an estimator of θ, Γ be the class of all maximal translation invariant functions of X and assume the loss in estimating θ by $\delta(X)$ is $k(\delta(X) - \theta)^2$. A necessary and sufficient condition that among all estimators of the form $\delta(X) + u\gamma(X)$, $\gamma \varepsilon \Gamma$, u constant, $\delta(X)$ uniquely minimize the risk is given and an explicit construction of the minimum risk estimator is derived therefrom. In the particular case where $\delta(X)$ has the translation property, the class of estimators of the form $\delta(X) + u\gamma(X)$ is the class of all estimators having the translation property. Thus, a construction of the minimum risk estimator having the translation property is exhibited of which the constructions given by Pitman (1939) and Blackwell and Girshick (1954) in the case where θ is a translation parameter are special cases. An example is given in which θ is not a translation parameter in the usual sense but estimators having the translation property are naturally admitted. Under an appropriate transformation the results are applicable to the estimation of scalar parameters. (Received July 2, 1958.)

59. The Moments of the Maximum of Partial Sums of Independent Random Variables. JOHN S. WHITE, Minneapolis-Honeywell Regulator Co.

Let $X_1, \cdots, X_n$ be independent identically distributed random variables. Let $S_k = \sum_{i=1}^{k} X_i$, $S_k^+ = \max(0, S_k)$, $\bar{S}_n = \max_{k \leq n}(S_k^+)$, $m_i(k) = E((S_k^+)^i)$ and $M_j(n) = E(\bar{S}_n^j)$. By successive differentiation of Spitzer's Theorem (*Trans. Amer. Math. Soc.*, Vol. 82, 1956) the following recursion relation for the moments of $\bar{S}_n$ is obtained:

$$E(\bar{S}_n^{j+1}) = M_{j+1}(n) = \sum_{k=1}^{n} \sum_{i=0}^{j} \binom{j}{i} [m_{i+1}(k)/k] M_{j-i}(n-k).$$

(Received July 2, 1958)

60. A Characterization of Triangular Association Scheme. S. S. SHRIKHANDE, University of North Carolina. (By title)

If a partially balanced design with two associate classes for $v = n(n-1)/2$ is triangular, (Bose and Shimamoto, *J. Amer. Stat. Assn.*, Vol. 47 (1952) pp. 151–190) then its parameters are given by $v = n(n-1)/2$, $n_1 = 2n - 4$, $p_{11}^1 = n - 2$, $p_{11}^2 = 4$. Connor (*Ann. Math. Stat.*, Vol. 29 (1958), pp. 262–266) has proved that if $n \geqq 9$, a design with above parameters is necessarily triangular. The following Lemma is established and it is utilized to prove that Connor's result is true for $n = 5, 6$ as well.

LEMMA: If for a design with above parameters, the 1-associates of any treatment x can be divided into two sets $(y_1, y_2, \cdots, y_{n-2})$, $(z_1, z_2, \cdots, z_{n-2})$ such that $(y_i, y_j) = (z_i, z_j) = (y_i, z_i) = 1$ and $(y_i, z_j) = 2$, $i \neq j = 1, 2, \cdots, n-2$, then the design is triangular. (Received July 2, 1958.)

61. A Problem in Two-Stage Experimentation. (Preliminary Report) DONALD L. RICHTER, University of North Carolina.

Let N_1 and N_2 be two normal populations with unknown variances and an unknown but common mean μ; it is desired to estimate μ using a fixed number n of observations. For this problem, a two-stage sampling procedure is proposed in which m observations are taken from each of N_1 and N_2 in the first stage and, depending on the observed values, $n - 2m$ observations are taken from one or the other population in the second stage. Associated with an estimator of μ, a risk function is defined which is equal to the variance of the estimator multiplied by a suitable stabilizing factor. For a particular unbiased estimator, it is shown that the minimax value of m is asymptotically equal to $cn^{2/3}$. Extensions in several directions are being studied. (Received July 2, 1958.)

62. Tests for the Validity of an Exponential Distribution of Life. BENJAMIN Epstein, Stanford University. (By title)

In this paper a number of procedures are given for testing, on the basis of life test data, whether there are substantial departures from an exponential distribution of life. The particular procedures that one should adopt depends on the class of alternatives one is testing against. A number of the tests are based in an essential way on fundamental properties of Poisson processes. (Received July 2, 1958.)

63. Stochastic Models for Length of Life. BENJAMIN EPSTEIN, Stanford University. (By title)

Various models for length of life are considered in this paper. Among these are (1) models which we call exponential (these involve Poisson processes and appropriate generalizations

of such processes); (2) models based on the conditional probability of failure function; (3) extreme value models. Implications of and interrelations among the various models are discussed. Many examples are given. As examples of models (1) and (2) one may cite the recent paper by Z. W. Birnbaum and S. C. Saunders (*J. Amer. Stat. Assn.*, Vol. 53 (1958), pp. 151-160) in which they give a statistical model for the life length of structures under dynamic loading (i.e., fatigue) and a recent report by George H. Weiss in which it is shown that some kinds of mechanical failure, such as creep failure of oriented polymenic filaments under tensile stresses, can be viewed as "pure death" processes. An example of where model (3) may be relevant is in phenomena involving corrosion. (Received July 2, 1958.)

64. Truncation and Tests of Hypotheses. OM P. AGGARWAL AND IRWIN GUTTMAN, Purdue University and Princeton University.

Consider a normal distribution with variance σ^2 and a sample from the distribution obtained from this normal distribution by truncating it at the same distance a on both sides of the mean. The distribution of the sample mean for sample sizes up to 4 is obtained explicitly and the results of applying the usual tests of hypotheses for one-sided testing of the mean of a normal distribution are examined when a and σ^2 are known. Some tables are given and it is found that the loss in power decreases very rapidly with the distance of the alternative value of the mean from the one tested and also with the distance of the truncation from the mean. (Received July 2, 1958.)

65. Mathematical Outline of Polyvariable Analysis (Including Random Balance). F. E. SATTERTHWAITE, Statistical Engineering Institute.

A polyvariable technique for statistical analysis is defined as any estimation procedure applied to the linear model, $Y = BZ + E = BZ + EI = AX$, $A = (B, E)$, $X = (Z, I)$, which gives estimates for all (or of some) of the A unknowns with associated confidence limits that are *valid* and *finite* without restrictions on the number of A unknowns in the model. Specifically the number of unknowns may exceed, and often will greatly exceed, the number of data sets. The theoretical minimum number of data sets is 2. The necessary minimum for a specific application to give useful precision depends primarily on the signal-noise ratio for the available data. In many types of applications satisfactory precisions are obtained with 5 to 30 data sets for models containing a large number of unknowns. This paper is a mathematical outline of method and justification (including, in most cases, formal proofs) for the more important classes of polyvariable methods: (1) Polygression, (2) Bigression, (3) Quadratic, (4) Homovariance, (5) Hetervariance, (6) Random Balance, (7) Split Data. (Received July 3, 1958.)

66. Statistical Theory of Some Quantal Response Models. ALLAN BIRNBAUM, Columbia University. (By title)

Let $V = (S_1, \cdots S_k)$, where S_g's are independent Bernoulli observations: Prob $\{S_g = 1\} = P_g(y)$, a known strictly-increasing function of the unknown real-valued parameter y, Prob $\{S_g = 0\} = Q_g(y) = 1 - P_g(y)$, for $g = 1, \cdots k$. If $P_g(y)$ depends on known parameters $a_g, b_g, \cdots$, whose values the experimenter may determine, these are called "design parameters." Fisher's (*Phil. Trans. Roy. Soc. London*, A, Vol. 222(1922), pp. 363-366) method in treating estimation and design problems in the dilution series model ($P_g(y) = 1 - \exp(-a_g y)$, $g = 1, \cdots k$) is formulated more explicitly, particularly the use of the practical equivalence of designs having similar information curves $I(y) = \sum I_g(y)$, where $I_g(y) = (\partial/\partial y P_g(y))^2/P_g(y)Q_g(y)$. The "information area" $\int I(y)\, dy$ is introduced and

used in various design problems. Point- and interval-estimation, hypothesis-testing, and other inference problems, and related problems of design and comparison of experiments, are treated, using efficient or simpler less efficient statistics, with examples from mental tests, industrial gauging, genetics and special analytical bioassays. It is shown that a necessary and sufficient condition for existence of a sufficient statistics is that, in terms of $z = z(y) = \log P_1(y)/Q_1(y)$, the model have the logistic form:

$$P_g(y) = (1 + \exp(-a_g z - b_g)^{-1} \text{ for } g = 1, \cdots k;$$

then $\sum a_g S_g$ is sufficient. (Invited address given at Los Angeles meeting, December, 1957. Received July 7, 1958.

67. Statistical Theory of Tests of a Mental Ability. ALLAN BIRNBAUM, Columbia University. (Invited paper)

Several writers (F. Lord, *Psychometrika*, Vol. 18(1953), pp. 57-76, and references therein) have studied the following model of a mental-ability test consisting of k items: Let $S_g = 1$ or 0 as a subject's response to item g is correct or not, $g = 1, \cdots k$. Then if a subject has ability y, the probability that his response pattern will be $V = (S_1, \cdots S_k)$ is

$$\Pi_{g=1}^{k} P_g(y)^{S_g} Q_g(y)^{1-S_g},$$

where $P_g(y) = \Phi(a_g y - b_g)$, $g = 1, \cdots k$, and $\Phi(u)$ is the standard normal c.d.f. Assuming item-parameters a_g, b_g known, problems of inference and design (choice of k, a_g's, b_g's) have been treated, as have Bayesean problems with y distributed according to $\Phi(y)$. Replacing $\Phi(u)$ by the logistic c.d.f. $\Psi(u) = (1 + \exp(-u))^{-1}$ gives a more tractable, perhaps equally valid, "logistic test model": $t = \sum a_g S_g$ is a sufficient statistic, typically nearly normal for each y; hence a design $(a_1, b_1; \cdots; a_k, b_k)$ is practically characterized by its "information curve" $I(y) = \partial/\partial y E(t \mid y) = \text{var}(t \mid y)$. If $I(y) \doteqdot c\Psi(ay - b)$ for some a, b, c (as in cases of principal interest), properties of Bayes estimates $E(y \mid t)$ are given as functionals of the c.d.f. of a weighted sum of two independent (Fisher's) z variables; numerical illustrations are given. A simple efficient method of estimating a_g's, b_g's is given. (Received July 7, 1958.)

68. On Logistic Order Statistics. ALLAN BIRNBAUM, Columbia University. (By title)

Plackett (*Ann. Math. Stat.*, Vol. 29(1958), pp. 131-142) has demonstrated the usefulness and tractability of logistic order statistics in treating problems involving order statistics from various distributions. The present more descriptive investigation of logistic order statistics, a by-product of development of statistical theory of a "logistic model" of ability tests, is a contribution to the comparative study of order statistics initiated by Hastings et al. (*Ann. Math. Stat.*, Vol. 18(1947), pp. 413-426). Because with suitable choice of scale parameter the logistic c.d.f. approximates the standard normal c.d.f. with error $< .01$, the logistic model is of interest, and may be sometimes preferred, when equally plausible, to the more usual (but less tractable, as regards order statistics) normal model of the population sampled. The presentation illustrates the effect on order statistics of such a change of parametric assumptions. Tables and graphs compare means and variances of logistic and normal order statistics for various sample sizes. The tractability of asymptotic variance and covariance formulae, and of some distributions related to extreme values, is illustrated. The distribution of each logistic order statistic coincides (to within a scale-factor) with a certain Fisher's—z distribution, for which extensive tables and approximation methods are available. (Received July 7, 1958.)

69. **Industrial Experience with Fractional Replicates.** CUTHBERT DANIEL, (Invited paper)

Typical conditions of industrial experimentation (including numbers of factors simultaneously studied, number and magnitude of effects sought, and restrictions on time and costs) are reviewed. For meeting these, the sequential use of nested sub-fractions of fractional replicate designs in the 2^{p-q} series is described. Since generally choice of a most informative initial sub-fraction is incompatible with choice of a most informative complete fractional replicate, the relative merits of each type, and of intermediate types, are discussed. A number of sequential designs are given. Methods are recommended and illustrated for inspection and criticism of data from 2^{p-q} experiments by using the graph (on appropriate probability paper) of the empirical c.d.f. of absolute values of contrasts, to detect one or two mavericks, inadvertent plot-splitting, antilognormal data, and the presence of several real effects. The distribution of this c.d.f. is studied under several hypotheses, and the use is described of the operating characteristics of a related statistic given by A. Birnbaum. Partial duplication is recommended when an unbiased estimate of error variance is required at an early stage. (Received July 7, 1958.)

70. **On the Analysis of Factorial Experiments without Replication.** ALLAN BIRNBAUM, Columbia University. (By title)

Inferences from factorial experiments without replication are usually based on a formal assumption that certain interactions are zero. In an altogether exploratory research situation, any statistical model giving a formal basis for informative inferences will typically be too schematic and restrictive of unknown conditions to be claimed "valid," or a basis for inferences which are "valid" except in the hypothetical formal sense; such a model is, perhaps along with other models, a basis for "plausible inferences," i.e., inferences drawn in a formally-valid manner, based on a model which is more or less plausible. Under some conditions (which are reviewed), the following schematic model is usefully plausible: The m contrasts a_i are independent, normal, homoscedastic; at most (any) r of their means are non-zero. For $r = 1$, to decide which, if any, mean is non-zero, the statistic $\max_i a_i^2/\sum_i a_i^2$ is optimal. An alternative graphical procedure developed by C. Daniel, which has important advantages, is related to the ratio of $\max_i |a_i|$ to another ordered $|a_i|$. Critical values, power and related properties, comparisons with more conventional statistics, and discussion of cases $r > 1$, are given. (Received July 7, 1958.)

71. **Linear Regression in the Multivariate Normal Case.** CHARLES STEIN, University of California, Berkeley.

The problem of estimating the regression vector of one random variable on p others when all have a joint normal distribution is considered. There are $n \geqslant p + 2$ observations on the whole vector, the mean is assumed 0 for simplicity, and the loss is taken to be the mean squared error of prediction when the estimated regression vector is used to make a prediction on the basis of a new random observation on the predictors, divided by the residual variance. The usual (maximum likelihood) estimate of the regression vector is minimax. It is admissible for $p = 1$, $n \geqslant 4$ and for $p = 2$ and n sufficiently large. For $p \geqslant 3$ it is intuitively clear (by analogy with the problem of estimating the mean of a multivariate normal distribution) that the usual estimate is not admissible. One possible method of improvement is to multiply the usual estimate by a constant depending on the population multiple correlation coefficient, which can be estimated from the sample coefficient. This will be more helpful if a guessed regression or a regression on a small selected set of predictors is first subtracted out. Other possible improvements are suggested when the covariance matrix of the predictors is known. It should also be possible to make further

improvements when the covariance matrix of the predictors is not known but guessed or estimated on the basis of an additional sample. (Received September 4, 1958).

72. **Some Population Estimation Models and Related Limit Distributions.** RONALD PYKE AND N. DONALD YLVISAKER, Stanford University.

The following two stage tag-and-sample model is studied. During stage I, $J + 1$ samples of sizes $m_0, m_1, \cdots, m_J$ are taken from a population of size S_1. In each sample, all untagged members are tagged and the sample replaced. During a later stage II, $K + 1$ samples of sizes $(n_0, n_1, \cdots, n_K)$ are taken in each of which all tagged members are tagged with a different tag than that used in stage I. The time interval between stages is assumed to be large relative to the time required to perform the tagging and sampling. Constant deterministic birth and death rates μ, and ρ are assumed during the intermediary time period. Maximum Likelihood estimates of S_1, μ and ρ are obtained under both Poisson and Binomial assumptions on the distribution of the recovery random variables (r.v.). Some general limit theorems are derived and applied to show that under a suitable reparametrization (corresponding to large sample and population sizes) the recovery r.v.'s and the Maximum Likelihood estimates are asymptotically normally distributed. A further generalization in which the sample sizes are assumed to be r.v.'s is considered. These results are then applied to data obtained from actual field experiments. (Received July 7, 1958.)

73. **Applications of a certain Representation of the Wishart Matrix.** ROBERT A. WIJSMAN, University of Illinois.

Let X be a $p \times n$ matrix ($p \leqq n$) whose columns are independent and distributed like $N(0, \Sigma)$. It is known (e.g., J. G. Mauldon, *J. Roy. Stat. Soc.*, Ser. B, Vol. 17 (1955) pp. 79–85) that the Wishart matrix XX' can be written as $CTT'C'$, where $CC' = \Sigma$, T is lower triangular with independent elements T_{ij}, T_{ii} is χ_{n-i+1} ($i = 1, \cdots, p$) and all T_{ij} with $i > j$ are $N(0, 1)$. This allows representation of any function of the Wishart matrix in terms of independent normal and χ variables. If the population correlation between two variates is ρ, the sample correlation r can be represented by $r/(1 - r^2)^{1/2} = (T_{21} + T_{11}\, \rho/(1 - \rho^2)^{1/2})/T_{22}$ (this representation was also obtained by G. Elfving, *Skand. Aktuarietids.*, Vol. 30 (1947), pp. 56–74). This can be described as a non-central t_{n-1} variable, with random non-centrality parameter $T_{11}\, \rho/(1 - \rho^2)^{1/2}$. If the population multiple correlation between one variate and the remaining $p - 1$ is $\bar{R}$, the sample multiple correlation R can be represented by $R^2/(1 - R^2) = ([T_{p1} + T_{11}\, \bar{R}/(1 - \bar{R}^2)^{1/2}]^2 + \sum_{i=2}^{p-1} T_{pi}^2)/T_{pp}^2$. This is a non-central $F_{p-1, n-p+1}$ variable, with random non-centrality parameter $T_{11}^2\, \bar{R}^2/(1 - \bar{R}^2)$. The sphericity criterion Z (T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*, Wiley, New York, 1958, section 10.7) in a bivariate population, when the hypothesis is true, can be represented by $Z/(1 - Z) = 2T_{11}T_{22}/((T_{11} - T_{22})^2 + T_{21}^2)$, which is an $F_{2n-2,2}$ variable. (Received July 7, 1958.)

74. **Order Statistics and Estimation.** M. M. RAO, University of Minnesota. (Introduced by Milton Sobel) (by title)

Let (1) $f(x) = e^{-x}$, if $x > 0$, and zero otherwise, and X_i be the ith order statistic from a sample of N independent observations from the population defined by (1). The following results are proved. (I): Let $1 \leqq r_1 < r_2 < \cdots < r_p \leqq N$ be a set of fixed integers and $X_{r_1}, X_{r_2}, \cdots, X_{r_p}$ be a p-set (subset) of the order statistics $X_1 < X_2 < \cdots < X_N$. Then the order statistics define a Stochastic Process with $r_1, r_2, \cdots$, as the parameter set, which has independent (but not stationary) increments. The finite-dimensional distributions of the process in terms of its log characteristic function are given by $\psi_{r_1, r_2, \ldots, r_p}(\xi_1, \xi_2, \cdots \xi_p)$

$= \log \varphi_{r_1 \cdot r_2, \dots, r_p}(\xi_1, \xi_2, \cdots \xi_p) = \log N!/(N - r_p)! - \sum_{j=0}^{p-1} \sum_{m=r_j}^{r_{j+1}-1} \log (N - i\eta_{j+1} - m)$ where $\eta_j = \sum_{l=j}^{p} \xi_l$, $r_0 = 0$, and φ is the ch.f. (II): Let X_{r_i} and r_i be defined as in (I). Then $\{X_{r_i}, 1 \leqq r_i \leqq N\}$ forms a Markov Process in the strict sense as well as in the wide sense (in either case, the Process is non-stationary). (III): Some problems of interest (in physiological data) are the following: (i) The r_i are random but N is fixed. Suppose Prob $\{r_k = i_k, k = 1, 2, \cdots, p \mid r_i < r_{i+1}, i = 1, 2, \cdots p - 1\} = p_{i_1 \cdot i_2, \dots, i_p}$, and $p_{i_1, \dots, i_p} \geqq 0$, $\sum_{i_1, \dots, i_p} p_{i_1, \dots, i_p} = 1$, where $(i, < i_2 < \cdots < i_p) = 1, 2, \cdots N$, $i_0 = 0$, and the p's depend on a set of constants, $(\lambda_1, \lambda_2, \cdots, \lambda_k)$. Then the X_{r_i} defined similar to those in (I) still form a Stochastic Process whose finite-dimensional d.f.'s are determined by the ch.f. $\varphi(\xi_1, \cdots, \xi_p; \lambda_1, \cdots \lambda_k) = \sum_{i_1, \dots, i_p} p_{i_1 \cdot i_2, \dots, i_p} [\prod_{m=1}^{i_p} (N - m + 1)/\prod_{j=1}^{p} \prod_{m=i_{j-1}}^{i_{j+1}} (N - i\eta_j + m)]$. (ii) The case when N is a random variable. Specifying the d.f.'s in some cases of interest for the r_i, the limit d.f.'s of some linear combinations of X_i are considered. The estimation of the constants (A, θ) and the distribution of the $(\hat{A}, \hat{\theta})$ are treated using the above results when in (1) x is replaced by $(x - A)/\theta$, $\theta > 0$. (Received July 7, 1958.)

75. A Note on Order Statistics and Stochastic Independence.

GERALD S. ROGERS, University of Arizona.

The following theorem is proved. Let x be a continuous or discrete type real random variable. Let $x_1 \leqq \cdots \leqq x_n$ be the order statistics based on a random sample of size n from this x distribution. Let $z = z(x_1, \cdots, x_j)$ be a statistic based on the first $j < n$ items only. If z is stochastically independent of x_j, then z is stochastically independent of all x_k, $j < k \leqq n$; if z is stochastically independent of some x_k, $j < k \leqq n$, then z is stochastically independent of x_j and hence of all x_k, $j \leqq k \leqq n$. The first result is direct, since in terms of the conditional probability density functions, $g(z \mid x_j) = g(z \mid x_j, \cdots, x_n)$. For the second part, in $g(x_1, \cdots, x_{k-1} \mid x_k)$, let x_k be considered as a "parameter." Then, $(x_{k-1} \mid x_k)$ is a "complete single sufficient statistic" for x_k; also, the distribution of $(z \mid x_k)$ is free of the "parameter x_k." By a well known theorem, (Basu, *Sankhya*, Vol. 15 (1955), pp. 377–380), $(z \mid x_k)$ and $(x_{k-1} \mid x_k)$ are stochastically independent. It follows that z and x_{k-1} are stochastically independent; similarly, with an induction, z and x_k, $j \leqq k \leqq n$, are stochastically independent. (Received July 7, 1958.)

76. A model for Failure Data and its Applications. (Preliminary report)

ANDRE G. LAURENT, Wayne State University.

When a "ageing process" takes place, the response pattern of a "system" to a stimulus X does not follow an exponential distribution. The model $S(t) = \exp [1 + t - \exp (t)]$, where $S(t)$ is the "survival function," i.e., the integral of the "X-to-failure" distribution and $t = (X - X_0)/\tau$, has been proposed to meet this situation and tables provided for its use (*Oper. Res.*, February, 1957, p. 150; *Oper. Res. 13th National Meeting*, p. 35.) The present paper describes the more important features of the model above and gives the formulas for the expected values and the covariance matrix of the order statistics of a sample of size n. Tables of the expected values and the variances for $n = 1$ to 15, of the covariances for $n = 2$ to 5 are provided. The minimum variance linear unbiased estimates of the parameters of the distribution based on order statistics are studied for small samples and compared to other estimates from the viewpoint of efficiency. Related models are considered. (Received July 7, 1958.)

77. A Convolutive Class of Monotone Likelihood Ratio Families.

S. G. GHURYE AND DAVID L. WALLACE, University of Chicago.

A one-dimensional family $f(x, \theta)$ of densities on the real line or of probabilities on the integers, with the real parameter θ, is called a monotone likelihood ratio family if the ratio

$f(x, \theta')/f(x, \theta)$ is nondecreasing in x for $\theta \leq \theta'$. If several monotone likelihood ratio families each have all probability on two points which are the same for all families and all parameter values, then their convolution is a monotone likelihood ratio family. The extent to which similar results hold for distributions on three and more points and, with appropriate extensions of definitions, for multidimensional distributions on the vertices of the simplex and the cube is determined. A sufficient condition that the convolution of monotone likelihood ratio families be a monotone likelihood ratio family is that for each family, the ratio $f(x + h, \theta)/f(x, \theta)$ be non-increasing in x for all $h > 0$. (Received July 7, 1958.)

78. **On the Exact Joint Distribution of the First Two Serial Correlation Coefficients.** V. K. MURTHY, University of North Carolina.

Any test of the hypothesis that up to a particular lag the true serial correlation coefficients are zero against some suitable alternative seems to necessitate knowledge of the joint distribution of serial correlation coefficients. As far as the author is aware even in the case of the first two serial correlation coefficients, the joint distribution has not so far been obtained in a simple closed form. In this note the joint distribution of r_1 and r_2 has been obtained for samples of independent normal variates assuming the sample size to be of the form $4n + 1$ where n is a positive integer and adopting the circular definition suggested by Hotelling. This result has been obtained using a result of R. L. Anderson on the characteristic roots of the serial covariance, and inversion formulae for the distribution of ratios of quadratic forms given by Gurland. Some properties of the joint distribution are obtained. The case of more than two serial correlation coefficients will be dealt with in a subsequent paper. (Research under ONR contract Nonr 855(06)). (Received July 7, 1958; revised July 28, 1958.)

79. **Confidence Bounds Associated with a Test for Symmetry.** R. GNANADESIKAN, The Procter and Gamble Company.

In a p-variate nonsingular normal distribution $N[\mu, \Sigma]$, one may be interested in testing a hypothesis of symmetry in the means, viz., that the p variates have the same mean. The tests obtained by using either the extended Type I union-intersection principle or the likelihood ratio are identical and it is well known that they are equivalent to an F-test with appropriate degrees of freedom. However, from the standpoint of confidence procedures, it is shown that the usual elliptical region can be replaced by simultaneous interval statements on parametric functions which are measures of departure from the null hypothesis. Also using a "truncation" procedure it is shown that one can study contrasts which are of particular interest and are components of the null hypothesis. The interval statements, which have a joint confidence coefficient $\geqq (1 - \alpha)$, are easier to use than the elliptical regions which have an exact confidence coefficient $(1 - \alpha)$. Received July 7, 1958.)

80. **On Stochastic Approximation.** C. DERMAN AND J. SACKS, Columbia University.

A very general theorem was proved by Dvoretzky ("On stochastic approximation", *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*) on the convergence, with probability one and in mean square, of stochastic approximation procedures. Wolfowitz ("On stochastic approximation methods," *Ann. Math. Stat.*, Vol. 27, 1956) presented a different proof. In this paper a third and simpler proof of the probability one convergence is given. Also, the probability one version is extended directly to the multi-dimensional case with absolute values of real numbers replaced by lengths of vectors. The one-dimensional theorem is a consequence of the following easily proved

lemma. *If* $\{a_n\}$, $\{b_n\}$, $\{c_n\}$, $\{\delta_n\}$ *and* $\{\xi_n\}$ *are sequences of real numbers satisfying the following conditions*: (i) $\{a_n\}$, $\{b_n\}$, $\{c_n\}$ *are positive*, (ii) $\{\xi_n\}$ *are non-negative*, (iii) $\lim a_n = 0$, $\sum b_n < \infty$, $\sum c_n = \infty$, $\sum \delta_n < \infty$, (iv) $\xi_{n+1} \leqq \max(a_n, (1 + b_n)\xi_n + \delta_n - c_n)$ *for all n greater than some N, then* $\lim \xi_n = 0$. The multi-dimensional theorem follows from a slightly modified version of the above lemma. (Received July 7, 1958.)

81. **A Classification Problem.** Oscar Wesler, University of Michigan.

The following version of "the problem of the k-faced die" is considered: Nature's pure strategies make up two sets of states, Ω_1 consisting of the $k!$ states got by permuting a known probability distribution $p = (p_1, p_2, \cdots, p_k)$ over the faces $1, 2, \cdots, k$ of a k-faced die, Ω_2 consisting similarly of the $k!$ states arising from a known distribution $q = (q_1, q_2, \cdots, q_k)$. Classification is made on the basis of N observations given by the sufficient statistic $r = (r_1, r_2, \cdots, r_k)$ representing the number of times each face appears. Let φ be a randomized statistical decision procedure, and let $\alpha(\varphi)$ and $\beta(\varphi)$ be the maxima of the probabilities of errors of the first and second kind, respectively. Then we wish to minimize $\beta(\varphi)$ subject to $\alpha(\varphi) = \alpha_0$. The class of unique symmetric procedures φ^* optimal in this extended Neyman-Pearson sense is found by a game-theoretic, minimax method, and from the invariance of the problem under the symmetric group of permutations on k letters. A simplification is given for large N, in which the φ^* are replaced by *kaleidoscopic* tests, determined by a one-parameter family of hyperplanes and their symmetric images. Finally, it is shown that, for $k = 2$, the φ^* and the kaleidoscopic approximations are in exact agreement for every N. (Received July 7, 1958.)

82. **Generalization of Palm's Loss Formula for Telephone Traffic.** V. E. Beneš, Bell Telephone Laboratories, Inc.

Let F be a real non-negative function on a space X, let $\mathfrak{F}$ be a Borel field of X-subsets, and let ξ_k, $k = 0, 1, 2, \cdots$ be a stationary Markov process taking values in X, with transition function $p(\xi, A)$ for ξ in X and A in $\mathfrak{F}$. We interpret the numbers $F(\xi_k)$ as the interarrival times of telephone calls at a trunk group. There are N trunks, lost calls are cleared, and holdingtimes of trunks are independent, with a negative exponential distribution of mean, γ^{-1}. We prove the following result: If P is the stationary probability measure of ξ_k, then the chance of loss (of finding all N trunks busy) is $\left[\sum_0^N \binom{N}{n} A_n\right]^{-1} A_0 P(X)$, with $A_N = I$ and $A_n = K_N[I - K_N]^{-1} \cdots K_{n+1}[I - K_{n+1}]^{-1}$, where K_n is the operator whose action on a measure μ is defined by $K_n \mu(A) = \int X \int A \exp\{-n\gamma F(\xi)\} p(\eta, d\xi)\mu(d\eta)$. Palm's formula applies to the case $X = (0, \infty)$, $F(\xi) \equiv \xi$, ξ_k independent. Our formula has the same algebraic form as Palm's, but the multiplicative constants have been replaced by operators. The inverses indicated in our formula exist under weak hypotheses. (Received July 14, 1958.)

83. **Factorial Analysis of Life-Tests.** Marvin Zelen, National Bureau of Standards.

Consider a factorial experiment involving the factors A and B having levels a and b respectively. Let a life-test experiment be planned such that n items are tested for each of the ab factorial combinations and the test is terminated when exactly r ($r \leqq n$) of the test items have failed. Assume that the underlying distribution of failures for the (i, j) factorial combination ($i = 1, 2, \cdots, a; j = 1, 2, \cdots, b$) is $p(x_{ij}) = \theta_{ij}^{-1} \exp[-(x_{ij} - A_{ij})\theta_{ij}^{-1}]$ for $x_{ij} \geqq A_{ij}$, where $\theta_{ij} = ma_ib_jc_{ij}$. Maximum likelihood estimates are found for the param-

eters m, a_i, b_j, and c_{ij}. Likelihood ratio tests are given for testing various hypotheses for these parameters as well as approximations for the small sample distribution of these tests. (Received July 14, 1958.)

84. **Unbiased Estimation for Functions of Location and Scale Parameters.** R. F. TATE.

Integral transform theory is employed to obtain unbiased estimators (which in many cases have the minimum variance property) for functions of a location parameter θ and/or a scale parameter σ. Applications are made to the gamma distribution with parameters considered together and separately, and to truncated distributions in general. A simple formula is presented for estimating any differentiable function of a single location parameter of truncation; no calculation of distributions or conditional expectations is required in order to find a minimum variance unbiased estimator. Special attention is paid throughout the paper to the estimation of the functions $P(X \varepsilon A \mid \theta)$, $P(X \varepsilon A \mid \sigma)$, and $P(X \varepsilon A \mid \theta, \sigma)$, where A is an arbitrary Borel set. (Received July 21, 1958; revised July 25, 1958.)

85. **Theory of Successive Two-Stage Sampling.** (Preliminary report) B. D. TIKKIWAL (By title)

The general theory of Univariate Sampling on Successive Occasions have been studied by the author [J. Ind. Soc. Agric. Stat., Vol. 8 (1956), pp 84–90] under a specified sampling scheme and correlation pattern. Here the sampling units selected for study on various occasions are completely enumerated. The present paper gives the best estimator and its variance under the same sampling scheme when each of the primary units (assumed of the same size M) are not completely enumerated but observed only on a sub-sample of size m. It is shown that the form of the best estimator is the same as in the univariate case, when the pattern of correlation is the same at both the stages. It is further noted, that, for an infinite population and $M = \infty$, the variance of the best estimator on the hth occasion is given by ϕ_h/n_h''. V in the notations of the above paper and where V is the variance of the simple two stage sampling mean when only one primary unit is selected on the hth occasion. (Received August 1, 1958.)

86. **Functions of Markov Chains** (Preliminary Report); MURRAY ROSENBLATT, Indiana University.

Let $\bar{X}_n$, $n = 0, 1, \cdots$ be a Markov Chain with initial distribution $w_i = P[\bar{X}_0 = i]$ and stationary transition probability matrix $P = (p_{ij})$ $i, j = 1, 2 \cdots$. Let $\bar{Y}_n = f(\bar{X}_n)$ and let S_α, $\alpha = 1, 2, \cdots$, be the sets of states of $\bar{X}_n$ that f collapses into states of $\bar{Y}_n$. Let class one consist of those sets of states into which one has access with positive probability from at most one set of states. Class two is the complementary class of states. A necessary and sufficient condition that $\bar{Y}_n$ be Markovian (for a fixed f), whatever the initial distribution w_i of the $\bar{X}_n$ process, is given as follows: (i) If S_α belongs to class two, $\sum_{j \varepsilon S_\alpha} p_{ij} p_{j \cdot S_\beta} = p_{i \cdot S_\alpha} C_{S_\alpha \cdot S_\beta}$ for all i, β. Here $p_{i \cdot S_\alpha} = \sum_{j \varepsilon S_\alpha} p_{ij}$. (ii) Given any sequence of sets of states $S_1, S_2, \cdots, S_n$ where S_1 is of class two and $S_2, \cdots, S_{n-1}$ of class one, $\sum_{j \varepsilon S_{n-1}} p_{ij}^{(n-1)} p_{j \cdot S_n} = p_{i \cdot S_1} c_{S_{n-1} \cdot S_n}^{(n-1)}$ for all i if there is positive probability of the path $S_1 \to S_2 \to \cdots \to S_n$. (Received September 8, 1958.)

NEWS AND NOTICES

Readers are invited to submit to the Secretary of The Institute news items of interest

Personal Items

Dr. Churchill Eisenhart has been granted a Rockefeller award "in recognition of outstanding public service," and will spend the coming academic year in England, engaged in research. Dr. Eisenhart, chief of the Statistical Engineering Laboratory, which he organized in 1946, will be based at the Research Techniques Unit of the London School of Economics and Political Science, where he will continue preparation of material for a unified treatment of the fundamentals of measurement theory and practices as related to the needs of the biological, physical, social and behavioral sciences.

Ira G. Spicer, formerly Project Leader of Technical Analysis at Minneapolis-Honeywell, has taken a position as Research Engineer with the Lockheed Missile Systems Division in Sunnyvale, California.

In August, Nelson M. Blachman will take a two-year leave of absence from his job at the Sylvania Electronic Defense Laboratory, Mountain View, California, to become a Scientific Liaison Officer at the Branch Office of the U. S. Office of Naval Research in London, England, where he will carry on liaison with European scientists in the field of electronics.

Alan T. James of the Division of Mathematical Statistics, Commonwealth Scientific and Industrial Research Organization, Australia, will be a Visiting Lecturer at Yale University for the academic year 1958–1959.

Alfred Lieberman, formerly with the Bureau of Ships, has now joined the staff of the Institute for Defense Analyses, Washington, D. C.

Herman Wold has accepted an invitation to serve as Visiting Professor during the academic year 1958–1959 at Columbia University, Economics Department, New York.

Margaret P. Martin has taken the position of Associate Professor of Preventive Medicine (Biostatistics) at the Upstate Medical Center of the State University of New York at Syracuse. She formerly held a similar position at Vanderbilt University.

Hian Liang Ang, Drs. Math. completed his work for his Master's degree in Statistics at the University of California at Berkeley in October, 1957, and continued his work toward a Ph.D. degree. He goes back to Indonesia in August, 1958, to resume his post as Lecturer of Mathematics at the University of Indonesia at Bandung.

Mr. Ulysses V. Ward was appointed Instructor of Mathematics at Howard University in September, 1957.

John E. Freund has recently been appointed Chairman of the Department of Mathematics of Arizona State College at Tempe (soon to be called Arizona State University.)

Patrick Billingsley has accepted a position as Assistant Professor in the Department of Statistics of the University of Chicago.

John W. Morse has resigned from position as Head of Economics at Keuka College, New York, to teach Statistics full-time as Assistant Professor, Economics at Hobart and William Smith Colleges, Geneva, New York, doing statistical consulting and developing inventions.

John W. Mayne, Director, Operational Research, Royal Canadian Navy has been posted by the Defense Research Board to the SHAPE Air Defense Technical Centre, The Hague, Netherlands, to be Chief of an Operational Research Section. He expects to be in Europe for about three years.

William E. Jaynes has accepted a position as an Assistant Professor of Industrial Psychology and Statistics, and Director of the Bureau of Industrial Testing and Institutional Research at the University of Omaha in Omaha, Nebraska.

Charles T. Lewis has recently accepted a position as operations analyst in the Operations Research Group at Convair in Fort Worth, Texas.

R. E. Beckwith has accepted a position as Senior Research Engineer with the California Institute of Technology Jet Propulsion Laboratory.

Frances Campbell Amemiya has resigned her position as Chairman of the Department and Professor of Mathematics at George Pepperdine College in Los Angeles. She is now Associate Professor of Mathematics at the California Western University.

Dr. John E. Walsh, formerly with the Military Operations Research Division of Lockheed Aircraft Corporation is now with the Operations Research Group of the System Development Corporation, 2400 Colorado Avenue, Santa Monica, California.

H. W. G. Deeks has been appointed Statistician in the War Office, Whitehall, London, S.W.1.

Dr. Om P. Aggarwal has returned to Purdue University as Associate Professor after spending a year at the University of Alberta (1956–57) as Visiting Associate Professor and another at the University of Saskatchewan. While in Canada, Professor Aggarwal was also a Fellow at the Summer Research Institute of the Canadian Mathematical Congress which is held every summer at Queen's University, Kingston, Ontario.

Ingram Olkin, on sabbatical leave from Michigan State University, will be at Stanford University for the academic year 1958–1959.

J. E. Morton is serving as Statistical Adviser to the UN Economic Commission for Asia and the Far East in Bangkok, Thailand; he is also giving a course at Chulaloregkorn University in Bangkok on Linear Programming.

On July 2, 1958, Alan H. Gepfert became Director of Statistical Research of the Chicago and North Western Railway Company. The major present job is to develop economic models by which to forecast revenues. Also concerned with application of sampling and regression analyses to cost-finding and general corporate planning. Mr. Gepfert was formerly a member of operations research group and faculty of Case Institute of Technology.

The Data Processing Division of International Business Machines Corpora-

tion completed its move to Westchester during July. Its new address is International Business Machines Corporation, Data Processing Division, 112 East Post Road, White Plains, New York.

Professor Herbert Solomon, Teachers College, Columbia University, is spending a sabbatical year at Stanford at Berkeley. His mailing address is Statistics Dept., Stanford University, Stanford, California.

Fellowship and Research Opportunities, National Academy of Sciences—National Research Council, Division of Mathematics

The Division of Mathematics calls attention to the fact that several foundations and offices offer financial support for research in mathematics during the year 1959–60. A number of fellowships will be made available, as well as opportunities for mathematicians to engage in basic research. A partial list, with comments, is given below.

1. *National Science Foundation.* The National Science Foundation sponsors various fellowship programs in the sciences, including mathematics.

Predoctoral fellowships are awarded annually at the First Year, Intermediate, and Terminal Year levels of graduate study. Applications for 1959–1960 will be available in October 1958 from the National Academy of Sciences-National Research Council until the closing date in early January 1959; Award date—March 16, 1959.

Science Faculty fellowships for college science teachers (including mathematics) who plan to continue teaching and wish to increase their competence as teachers are at the present time offered semi-annually. Eligibility requirements include a baccalaureate degree and three (3) years of full-time experience in teaching natural science subjects at the collegiate level. Awarded annually. The program will be open from May to October. Awards will be announced in early December. Address all inquiries for information and applications to National Science Foundation, Division of Scientific Personnel and Education, Washington 25, D. C.

Postdoctoral fellowships (in making inquiry about postdoctoral awards specify program)

(1) *Regular* postdoctoral fellowships—primarily for recipients of the doctoral degree; awarded semi-annually. Program for 1959–60 concurrent with predoctoral program (see above) except that program closes in December. Information and applications will be available from the NAS-NRC. The program will also be open from July to early September 1959. Awards are announced in March and October.

(2) *Senior* postdoctoral fellowships—are open to persons who have held a doctoral degree in one of the basic fields of science for a minimum of five (5) years at time of application, or who have had equivalent training and experience. Awarded annually. Applications are available from the National Science Foundation, Division of Scientific Personnel and Education, Washington 25, D. C. The program will be open from May to October. Awards will be announced in early December:

Research Grants. The National Science Foundation also supports basic research in the mathematical sciences by means of grants. While proposals for such support are accepted at any time, individuals desiring support to begin in the summer or at the beginning of a fall semester should preferably submit their proposals in the mathematical sciences by November 1; persons desiring support to begin in the spring semester should preferably submit their proposals by May 1. Instructions for the preparation of proposals, contained in a booklet entitled Grants for Scientific Research, may be obtained upon request from the Program Director for Mathematical Sciences, National Science Foundation, Washington 25, D. C.

2. *Office of Naval Research.* The Office of Naval Research, through contracts with universities and other organizations, supports basic research in broadly selected fields of mathematics. Proposals should be directed to the Mathematics Branch, Office of Naval Research, Washington 25, D. C. In addition, postdoctoral research associateships in pure mathematics are being established under contracts with the ONR at selected universities. For details and application forms write to the above address.

3. *Air Force Office of Scientific Research.* The Air Force Office of Scientific Research supports research in mathematics directly through contracts with colleges, universities, foundations and industrial laboratories. Such organizations are encouraged to submit proposals for research in mathematical fields in which they specialize. Proposals should be mailed to the Commander, Air Force Office of Scientific Research, Attn: Mathematics Division, Washington, 25, D. C.

4. *Office of Ordnance Research, U. S. Army.* Among the functions of the Office of Ordnance Research is the support of basic research in mathematics. Proposals for projects are ordinarily made by individual scientists or groups of scientists in a form which leads to a contract between the Office of Ordnance Research and a university or research laboratory. For further information write to Commanding Officer, Office of Ordnance Research, Box CM, Duke Station, Durham, North Carolina.

5. *Fulbright Awards—Public Law 584 (79th Congress).* Approximately 400 awards are offered annually for university lecturing and postdoctoral research in all academic fields in Argentina, Australia, Brazil, Burma, Chile, Colombia, Ecuador, India, New Zealand, Pakistan, Paraguay, Peru, the Philippines and Thailand (competition for the preceding countries closes April 15, 1959); Austria, Belgium-Luxembourg, Republic of China, Denmark, Finland, France, Germany, Greece, Iceland, Iran, Ireland, Israel, Italy, Japan, the Netherlands, Norway, Turkey, and the United Kingdom including colonial dependencies (competition for the latter countries closes October 1, 1959). In both cases awards are for the academic year 1960-61 (the 1959-60 competition for Europe closes October 1, 1958), but in the former group of countries the academic year begins in the spring or summer instead of the autum. Awards are payable in foreign currency and usually include travel for the grantee, but not for members of his family, and a maintenance allowance, which may be adjusted in relation to the number of accompanying dependents up to four. Requests for information should be addressed to the Committee on International Exchange of Persons, Conference Board of Associated Research Councils, 2101 Constitution Avenue, Washington 25, D. C.

6. *National Bureau of Standards. Naval Research Laboratory. Air Research and Development Command.* Postdoctoral resident research associateships are available in a variety of sciences including mathematics and are tenable at the Washington, D. C. and Boulder, Colorado laboratories of the National Bureau of Standards; at the Naval Research Laboratory in Washington, D. C.; and at selected development and research centers of the Air Research and Development Command. Necessary facilities and equipment incident to the research of the associate will be provided. For further information write to Fellowship Office, National Academy of Sciences-National Research Council, 2101 Constitution Avenue, Washington 25, D. C. Applications for the 1959-60 program must be filed on or before January 19, 1959.

7. *Atomic Energy Commission.* The Division of Research of the Atomic Energy Commission through contracts with universities and other organizations supports research in the fields of numerical analysis, digital computer design, programming research, and related topics. Proposals should be submitted to the Division of Research, Atomic Energy Commission, Washington 25, D. C.

Brookhaven National Laboratory. Brookhaven National Laboratory, operated by Associated Universities, Inc. under contract with the Atomic Energy Commission offers postdoctoral research appointments in mathematics. Appointments are for one year, and may be renewed for one additional year. U. S. citizenship is not required, although Atomic

Energy Commission approval is a prerequisite. The appointee may work in numerical analysis, digital computing, mathematical physics, differential equations, probability and statistics, and various specialized branches including reactor theory, hydrodynamics, and orbit theory. Computational facilities are available. Letters from candidates should give details of personal history, scientific background, and qualifications; two letters of recommendation, one from the applicant's research professor, are required. Applications should be directed to M. E. Rose, Head, Applied Mathematics Division, Brookhaven National Laboratory, Upton, Long Island, New York.

September 1, 1958

S. S. WILKS, *Chairman*
Division of Mathematics
M. H. MARTIN, *Executive Secretary*
Division of Mathematics

Committee on Statistics

A new committee, the Committee on Statistics, has been established in the Division of Mathematics, NAS-NRC. It has been established in the Division as the successor to the Committee on Applied Mathematical Statistics which was appointed in 1942 and placed directly under the Academy-Research Council Governing Board. The funds in the custody of the earlier committee, and amounting to approximately $5,000, have been transferred to the custody of the new committee.

Cost-Free Digital Computer Time

As announced in the March, 1958, *Annals*, pp. 343–347, the Committee on Mathematical Tables of the Institute of Mathematical Statistics has made a survey of cost-free time on digital computers in the United States. The survey, in which 171 digital computer installations were queried, is now complete. Cost-free time is available at approximately 40 installations in at least 18 states in all parts of the United States. Members of the Institute of Mathematical Statistics who wish to avail themselves of some of this cost-free time to compute on a problem of general interest, i.e., a problem which might lead to publication of results in a professional journal, are invited to get in touch with the Chairman of the Subcommittee on Cost-Free Machine Time, Professor Fred C. Leone, Statistical Laboratory, Case Institute of Technology, Cleveland 6, Ohio. Advice on the preparation of specific tables (but *not* advice on programming or numerical analysis) is available from the other subcommittees and the reader is referred to the March, 1958, *Annals* for a complete listing of them.

D. B. OWEN, *Chairman*
Committee on Mathematical Tables

Preliminary Actuarial Examinations Prize Awards

The winners of the prize awards offered by the Society of Actuaries to the nine undergraduates ranking highest on the score of Part 2 of the 1958 Preliminary Actuarial Examination are as follows:

First Prize of $200: Daniel G. Quillen, Harvard University.

Additional Prizes of $100 each: Edward J. Barbeau, Jr., Toronto University; William H. Blake, Jr., George Washington University; Theodore M. Jungreis, Rensselaer Polytechnic Institute; David H. Krantz, Yale University; Joe Lipman, Toronto University; Dennis W. Moore, Harvard University; Theodore S. Rosky, State University of Iowa; Lawrence A. Shepp, Brooklyn Polytechnic Institute.

The Society of Actuaries has authorized a similar set of nine prizes for the 1959 examinations on Part 2.

The Preliminary Actuarial Examinations consist of the following three examinations: Part 1. Language Aptitude Examination. (Reading comprehension, meaning of words and word relationships, antonyms, and verbal reasoning.) Part 2. General Mathematics Examination. (Algebra, trigonometry, coordinate geometry, differential and integral calculus.) Part 3. Special Mathematics Examination. (Probability and statistics.)

The 1959 Preliminary Actuarial Examinations will be prepared by the Educational Testing Service under the direction of a committee of actuaries and mathematicians and will be administered by the Society of Actuaries at centers throughout the United States and Canada on May 13, 1959. The closing date for applications is April 1, 1959.

Postdoctoral Study in Statistics

Awards for study in statistics by persons whose primary field is not statistics but one of the physical, biological, or social sciences to which statistics can be applied are offered by the Department of Statistics of the University of Chicago. The awards range from $3,600 to $5,000 on the basis of an eleven month residence. The closing date for application for the academic year 1959–60 is February 16, 1959. Further information may be obtained from the Department of Statistics, Eckhart Hall, University of Chicago, Chicago 37, Ilinois.

Nonparametric Statistics

A revision is being made of "Bibliography of Nonparametric Statistics and Related Topics," *Journal of the American Statistical Association* 48 (1953) pp. 844–906. Material through 1959 is to be included with more emphasis, it is

hoped, on applications than previously. References (particularly to the non-English literature), reprints, and technical reports on the theory or applications of nonparametric statistics would be greatly appreciated. Also, corrections and additions to the original bibliography are desired.

University of Michigan Summer Program in Health Statistics

The School of Public Health, University of Michigan, will have a summer program in health statistics, June 18 through August 1, 1959. The faculty will be assembled from many of the schools of public health, and from the ranks of leading workers in the field of statistics in the health sciences. Tentative course titles are: Statistical Methods in Public Health, Management of Health Agency Records, Registration and Vital Statistics, Biostatistics in the Health Sciences, Demographic Methods in Public Health, Statistical Methods in Epidemiology, Sampling Techniques in the Health Sciences, Advanced Biostatistics in the Health Sciences, Statistical Methods in Biological Assay.

Further information can be obtained from F. D. Hemphill, School of Public Health, University of Michigan, Ann Arbor, Michigan.

Additional Doctoral Dissertations

The following doctorates, conferred in 1957, should be added to the list published in the June, 1958 issue of these *Annals*.

Kupperman, Morton, The George Washington University, major in mathematical statistics, "Further Applications of Information Theory to Multivariate Analysis and Statistical Inference."

McCall, Chester H., Jr., The George Washington University, major in mathematical statistics, "The Linear Hypothesis, Information, and the Analysis of Variance."

NaNagara, Prasert, Cornell, major in statistics, "Lattice Rectangle Designs."

New Members

The following persons have been elected to membership in The Institute

May 14, 1958, to June 23, 1958

Allen, (Rev.) Raymond W., Ph.D. (St. Louis University), Chairman, *Department of Mathematics, Xavier University, Cincinnati 7, Ohio.*

Amster, Sigmund J., M.S. (Columbia University), Student, University of North Carolina, *119 Harvey Street, Philadelphia, Pa.*

Burton, Ellison Stanley, B.A. (Amherst College), Systems Engineer (Statistician) Consultant, Harper Engineering Company, Santa Monica, California, *4650 East 19th Street, Tucson, Arizona.*

Cook, William H., A.B. (Hofstra College), Mathematical Statistician, *U. S. Bureau of the Census, Statistical Research Division, Washington 25, D. C.*

Edwards, Bernard, B.Sc. (London), Lecturer in Statistics, *Municipal College of Commerce, College Street, Newcastle Upon Tyne, Northumberland, England.*

Friedman, Morton Philbert, M.A. (Ohio State University), Student, Ohio State University, Columbus, Ohio, *865 Northwest Blvd., Columbus 12, Ohio.*

Goodman, Arnold F., B.S. (N. C. State College), Graduate Assistant, Stanford University, Stanford, California, *621 Harvard Avenue, Menlo Park, California.*

Hakim, Muhammad A., M.Sc. (Univ. of Calcutta) Graduate Student, *University of California, Department of Statistics, Berkeley 4, California.*

Harkness, William L., M.A. (Michigan State University) Special Graduate Research Assistant, *Department of Statistics, Michigan State University, East Lansing, Michigan.*

Higa, Seiko, B.A., (Pacific University) Statistician, *Finance Department, U. S. Civil Administration of the Ryukyu Islands, Naha, Okinawa, Ryukyu Islands.*

Hill, Bruce M., M.S. (Stanford University), Assistant in Statistics, Graduate Student, *Stanford Statistics Department, Stanford University, Stanford, California.*

Johnson, Whitney Larsen, M.S. (University of Minnesota), Instructor, Dept. of Math. Institute of Technology, University of Minnesota, Minneapolis 14, Minn. *2175A Folwell, St. Paul 8, Minn.*

Kaller, Cecil, M.A. (Saskatchewan), Research Fellow, Purdue University, Lafayette, Indiana, *Statistical Laboratory, Purdue University, Lafayette, Indiana.*

Masuyama, Motosaburo, (Doctor of Science), Chief of the Laboratory of Environmental Hygiene, Meteorological Research Institute, Tokyo, *Institute of Physical Therapy & Internal Medicine, Faculty of Medicine, Tokyo University, Bunkyo-ku, Tokyo, Japan.*

Middleton, David, Ph.D. (Howard University), Consulting Physicist, *23 Park Lane, Concord, Mass.*

Mills, Harlan Duncan, Ph.D. (Iowa State) Research Associate, Princeton University, Princeton, New Jersey, *186 Elm Road, Princeton, New Jersey.*

Nemenyi, Peter B., M.A. (Princeton University), Assistant Statistical Analyst, *Metropolitan Life Insurance Company, Madison Avenue, New York 10, N. Y.*; also Lecturer, Hunter College, Pk. Ave. and 68th Street, New York.

Niedzielski, Edmunk L., Ph.D. (Fordham University), Research Chemist, *E. I. DuPont De Nemours and Co., Petroleum Laboratory, P. O. Box 1671, Wilmington 99, Del.*

Okamoto, Masashi, M.S. (Tokyo University), Lecturer in Mathematical Statistics, Osaka University, Japan, *Nakanoshim, Kita-ku, Osaka, Japan.*

Pincus, Louis, B.S., (The City College of New York), Senior Statistician, New York City Department of Health, 125 Worth Street, New York 12, N. Y., *451 Kingston Avenue B-7, Brooklyn 25, New York.*

Randels, Robert B., Ph.D. (University of Michigan) Physicist, *Corning Glass Works, Houghton Park, Corning, New York.*

Robinson, Enders A., Ph.D., (Massachusetts Institute of Technology), Assistant Professor of Statistics, Michigan State University, *Department of Statistics Michigan State University, East Lansing, Michigan.*

Sato, Sokuro, (Tokyo Technical College) Assistant Professor, Faculty of Education, Saga University, Saga City, Saga Prefecture, Japan, *Dokkoo-koji, Mizugaemachi, Saga City, Saga City, Japan.*

Slud, Maurice H., M.A. (Columbia University), Mathematical Analyst, *General Electric Company, Missile and Ordnance Systems Department, 3198 Chestnut Street, Philadelphia 4, Pennsylvania.*

Smith, William Roger, M.S. (University of Wisconsin), Student, University of California, *5345 Zara Avenue, Richmond, California.*

Tsutakawa, Robert K., M.S. (University of Chicago), Quality Analyst A, Quality Control

Department, Pilotless Aircraft Div., Boeing Airplane Co., Seattle, Washington, *1103½ 3rd Avenue, Seattle, Washington.*

Welch, Peter D., M.S. (University of Wisconsin), Staff Engineer, *IBM Research Center, Yorktown Heights, New York.*

Zacks, Shelemyahu, B.A. (Hebrew University), *Statistician of the Building Research Station, Technion, The Technion Research Institute, Haifa, Israel.*

Zehna, Peter W., M.A. (University of Kansas) Research Assistant, Stanford University, *Applied Mathematics and Statistics Laboratory, Stanford University, Stanford, California.*

REPORT OF THE CAMBRIDGE, MASSACHUSETTS MEETING OF THE INSTITUTE OF MATHEMATICAL STATISTICS

The seventy-eighth meeting of The Institute of Mathematical Statistics and the twenty-first annual meeting was held at the Massachusetts Institute of Technology, Cambridge, Massachusetts, on August 25–28, 1958, in conjunction with the meetings of the American Mathematical Society, the Mathematical Association of America, the Society for Industrial and Applied Mathematics, and the Econometric Society.

The program of the meeting was as follows:

MONDAY, AUGUST 25, 1958

9:00 A.M. Invited Papers on Regression and Analysis of Variance

Chairman: FRANKLIN A. GRAYBILL, Oklahoma State University

1. *Variance Component Analysis in Models Where Effects Are Time Variables*, A. W. WORTHAM, (presented by LEROY FOLKS) Texas Instruments Inc., Dallas
2. *Industrial Experience with 2^{p-q} Fractional Factorial Experiments*, CUTHBERT DANIEL, New York City
3. *Confidence and Significance Procedures for Non-linear Models*, M. B. WILK, Bell Telephone Laboratories, Murray Hill

11:15 A.M. Wald Lecture I

Chairman: J. L. HODGES, JR., University of California, Berkeley

The Mathematical Basis of Fiducial Inference, JOHN W. TUKEY, Princeton University

2:00 P.M. Invited Papers on Estimation and Testing

Chairman: DONALD L. BURKHOLDER, University of Illinois

1. *Power of the Chi-square Test*, J. L. HODGES, JR., University of California, Berkeley
2. *On Solutions of Dorfman's Mass-Testing Problem*, MILTON SOBEL, Bell Telephone Laboratories, Allentown
3. *Linear Regression in the Multivariate Normal Case*, CHARLES STEIN, University of California, Berkeley

4:00 P.M. Invited Papers on Testing

Chairman: D. B. OWEN, Sandia Corporation

1. *Partial Orderings of Probabilities of Rank Orders*, I. RICHARD SAVAGE, University of Minnesota
2. *Simple Methods for Analysis of Two-Action Problems with Linear Costs*, ROBERT SCHLAIFER, Harvard University

8:00 P.M. 1958 Council Meeting

TUESDAY, AUGUST 26, 1958

9:00 A.M. Contributed Papers, I. (Simultaneous with Contributed Papers II)

Chairman: CHARLES STEIN, University of California, Berkeley

1. *Truncation and Tests of Hypotheses*, OM P. AGGARWAL, Purdue University AND IRWIN GUTTMAN, Princeton University (by title)
2. *Admissible Estimates and Maximum Likelihood Estimates (A Sketch of a Unified Theory of Estimation)*, ALLAN BIRNBAUM, Columbia University
3. *A Note on Estimating Translation and Scalar Parameters*, JOSEPH A. DUBAY, University of Oregon
4. *A Convolutive Class of Monotone Likelihood Ratio Families*, S. G. GHURYE AND DAVID L. WALLACE, University of Chicago
5. *Power and Control of Size of Some Optimal Welch-type Statistics*, ROGER S. McCULLOUGH AND JOHN GURLAND, Iowa State College
6. *Some Population Estimation Models and Related Limit Distributions*, RONALD PYKE AND N. DONALD YLVISAKER, Stanford University
7. *On the Choice of Sample Size in the Kolmogorov-Smirnov Tests*, JUDAH ROSENBLATT, Purdue University
8. *Theory of Successive Two-Stage Sampling*, (preliminary report) B. D. TIKKIWAL, Karnatak University, Dharwar, India (by title)
9. *On the Existence of Wald's Sequential Test*, ROBERT A. WIJSMAN, University of Illinois (by title)
10. *Functions of Markov Chains*, (preliminary report) MURRAY ROSENBLATT, Indiana University (by title)

9:00 A.M. Contributed Papers, II

Chairman: JACK NADLER

1. *Confidence Bounds Associated with a Test for Symmetry*, R. GNANADESIKAN, The Procter & Gamble Company (by title)
2. *Determining Bounds on Integrals with Applications to Cataloging Problems*, BERNARD HARRIS, George Washington University (by title)
3. *On the Exact Joint Distribution of the First Two Serial Correlation Coefficients*, V. K. MURTHY, University of North Carolina
4. *A Classification Problem Involving Multinomials*, OSCAR WESLER, University of Michigan
5. *Applications of a Certain Representation of the Wishart Matrix*, ROBERT A. WIJSMAN, University of Illinois
6. *On the Problem of Incomplete Data*, JUNJIRO OGAWA AND BERNARD S. PASTERNACK, University of North Carolina
7. *Unbiased Estimation for Functions of Location and Scale Parameters*, R. F. TATE, University of Washington (by title)
8. *Uniqueness of the L_2 Association Scheme*, S. S. SHRIKHANDE, University of North Carolina
9. *On the Asymptotic Minimax Character of the Sample d.f. of Vector Chance Variables*, J. KIEFER AND J. WOLFOWITZ, Cornell University, (by title)

11:15 A.M. Wald Lecture II

Chairman: WILLIAM H. KRUSKAL, University of Chicago

The Mathematical Basis of Fiducial Inference, (continued) JOHN W. TUKEY, Princeton University

2:00 P.M. Invited Papers on Probability and Stochastic Processes, I

Chairman: J. R. BLUM, University of Indiana

1. *A Moment-Problem with Restriction on Smoothness*, C. L. MALLOWS, Princeton University
2. *About the Central Limit Problem*, MICHEL LOÈVE, University of California, Berkeley
3. *Hausdorff Dimension and Information Theory*, PATRICK BILLINGSLEY, Princeton University and University of Chicago
4. *A Geometry of Binary Sequences Associated with a Class of Error-Correcting Codes*, ROY R. KUEBLER, JR., University of North Carolina

4:00 P.M. Special Invited Address

Chairman: SHANTI S. GUPTA, University of Alberta

Multiple Decision Selection Procedures, MILTON SOBEL, Bell Telephone Laboratories, Allentown

WEDNESDAY, AUGUST 27, 1958

9:00 A.M. Invited Papers on Sequential Analysis

Chairman: KENNETH J. ARNOLD, Michigan State University

1. *A Modification of Sequential Analysis to Reduce the Sample Size*, T. W. ANDERSON, Center for Advanced Study in Behavioral Sciences and Columbia University
2. *Binomial Sequential Testing*, COLIN R. BLYTH, Stanford University
3. *Unbiased Sequential Estimation for Binomial Populations*, MORRIS H. DEGROOT, Carnegie Institute of Technology

11:15 A.M. Wald Lecture III

Chairman: MARVIN ZELEN, National Bureau of Standards

The Interpretation of Fiducial Inference, JOHN W. TUKEY, Princeton University

2:00 P.M. Invited Papers on Probability and Stochastic Processes II. (Simultaneous with Invited Papers on Random Balance.)

Chairman: MAX WOODBURY, New York University

1. *Semigroups of Operators and Stochastic Processes*, A. V. BALAKRISHNAN, University of California, Los Angeles
2. *Independent Polynomials in Normal Variates*, R. G. LAHA, Catholic University of America and Columbia University
3. *On Multi-event Renewal Processes*, RONALD PYKE, Stanford University

2:00 P.M. Invited Papers on Random Balance

Chairman: FRANK J. ANSCOMBE, Princeton University

1. *Introductory Remarks*, FRANK J. ANSCOMBE, Princeton University
2. *On the Analysis of Screening Experiments*, E. M. L. BEALE, Princeton University, AND C. L. MALLOWS, Princeton University
3. *Analysis Methods for Randomly Balanced Factorial Designs*, A. P. DEMPSTER, Bell Telephone Laboratories and Harvard University
4. *Mathematical Outline of Polyvariable Analysis* (including *Random Balance*), F. E. SATTERTHWAITE, Statistical Engineering Institute, Wellesley Hills

4:00 P.M. Wald Lecture IV

Chairman: M. B. WILK, Bell Telephone Laboratories

What Importance Should We Place on Fiducial Inference? JOHN W. TUKEY, Princeton University

5:30 P.M. Business Meeting

8:00 P.M. 1959 Council Meeting

THURSDAY, AUGUST 28, 1958

9:00 A.M. Contributed Papers III (Simultaneous with Contributed Papers IV)

Chairman: ROBERT A. WIJSMAN, University of Illinois

1. *On a Limiting Distribution Due to Renyi*, D. G. CHAPMAN, University of Washington
2. *Stochastic Models for the Electron Multiplier Tube*, EDWARD K. DALTON, WILLARD D. JAMES, AND HOWARD G. TUCKER, University of California at Riverside
3. *On Stochastic Approximation*, C. DERMAN AND J. SACKS, Columbia University
4. *Single Server Queuing Processes with a Finite Number of Sources*, GERALD HARRISON, The Teleregister Corporatior
5. *A Model for Failure Data and its Applications*, ANDRE G. LAURENT, Wayne State University
6. *Generalization of Palm's Loss Formula for Telephone Traffic*, V. E. BENES, Bell Telephone Laboratories, Murray Hill
7. *The Moments of the Maximum of Partial Sums of Independent Random Variables*, JOHN S. WHITE, Minneapolis-Honeywell Regulator Co.
8. *Stochastic Models for Length of Life*, BENJAMIN EPSTEIN, Wayne State University and Stanford University (By title)
9. *Tests for the Validity of an Exponential Distribution of Life*, BENJAMIN EPSTEIN, Wayne State and Stanford Universities, (by title)

9:00 A.M. Contributed Papers IV

Chairman: M. V. JOHNS, JR., Stanford University

1. *Estimation of the Medians for Dependent Variables*, OLIVE JEAN DUNN, University of California, Los Angeles
2. *The Use of Sample Quasi-Ranges in Estimating Population Standard Deviation*, H. LEON HARTER, Wright Air Development Center
3. *Order Statistics and Estimation*, M. M. RAO, University of Minnesota (by title)
4. *A Note on Order Statistics and Stochastic Independence*, GERALD S. ROGERS, University of Arizona
5. *Aids for Fitting the Pearson Type III Curve by Maximum Likelihood*, J. ARTHUR GREENWOOD, Iowa State College AND DAVID DURAND, Massachusetts Institute of Technology
6. *On the Relationship Algebra and the Association Algebra of the Partially Balanced Incomplete Block Design*, JUNJIRO OGAWA, University of North Carolina (by title)
7. *Factorial Analysis of Life-Tests*, MARVIN ZELEN, National Bureau of Standards
8. *On the Analysis of Factorial Experiments without Replication*, ALLAN BIRNBAUM, Columbia University (by title)
9. *On Logistic Order Statistics*, ALLAN BIRNBAUM, Columbia University (by title)
10. *Statistical Theory of Some Quantal Response Models*, ALLAN BIRNBAUM, Columbia University (by title)
11. *Optimum Designs in Regression Problems*, J. KIEFER AND J. WOLFOWITZ, Cornell University, (by title)
12. *On the Bounds for the Variance of Mann-Whitney Statistic*, JAGDISH SHARAN RUSTAGI, Michigan State University, (by title)
13. *A Characterization of Triangular Association Scheme*, S. S. SHRIKHANDE, University of North Carolina (by title)
14. *A Problem in Two-Stage Experimentation*, DONALD L. RICHTER, University of North Carolina (by title)

11:15 A.M. Special Invited Address

Chairman: SAMUEL W. GREENHOUSE, National Institutes of Health

Estimation Methods in Multivariate Analysis, EVAN J. WILLIAMS, North Carolina State College

2:00 P.M. Special Invited Address

Chairman: MORRIS H. HANSEN, Bureau of the Census

On a Formal Structure of Professional Practice in Sampling, W. EDWARDS DEMING, New York University

3:15 P.M. Invited Papers on Mixed Topics

Chairman: H. A. DAVID, Virginia Polytechnic Institute

1. *The Number of Occupied Cells of a Particular Subclass (When Objects are Assigned to Cells at Random)*, HOWARD L. JONES, Illinois Bell Telephone Company, Chicago
2. *Statistical Theory of Tests of a Mental Ability*, ALLAN BIRNBAUM, Columbia University
3. *Properties of Some Control Chart Tests for Detecting Shifts in a Process Average*, S. W. ROBERTS, Bell Telephone Laboratories, New York.

REPORT OF THE PRESIDENT FOR 1958

To the membership of the IMS
Dear Friends:

I am writing to you shortly before my departure from the country, because I shall unfortunately not be here to address you personally at the annual meeting.

The presidency of the Institute is an honor greater than I had expected to receive in my whole life. I thank you all for your expression of recognition and confidence. I also thank for myself, and for all of us, the many officers, committee members, and representatives who have so loyally and competently handled the Institute's increasingly complex and serious business.

The president's job is not among the harder ones in the Institute, but I have found it interesting and instructive. The office entails some decisions and offers opportunities to make suggestions throughout the Institute from a central vantage point. It also offers opportunities to make mistakes, and I have made some. Those that have thus far come to light are mostly small and rectifiable.

Since the meeting at Atlantic City a year ago there have been regional meetings at Los Angeles, Gatlinburg, and Ames.

T. E. Harris announced a year ago that he wanted to relinquish the Editorship at the expiration of his term on July 1, 1958. An ad hoc committee consisting of William Cochran (Chairman), T. W. Anderson, M. S. Bartlett, T. E. Harris, W. A. Wallis, and S. S. Wilks recommended to the Council that W. H. Kruskal be appointed to the Editorship. The Council has accepted this recommendation and Kruskal has accepted the appointment. The *Annals* continues to grow and improve by leaps and bounds, as I expect you will hear in detail from the retiring Editor.

The Council decided in Atlantic City that, to meet the rising costs of printing a larger *Annals* at higher rates for printing, we ought to apply to the National Science Foundation for a grant of money for a three-year period. I hope

that it will be possible to announce along with the reading of this letter that the grant has been made.

The Council has before it a plan, submitted by an ad hoc committee, to approach the National Science Foundation for aid in preparing translations of Russian statistical and probabilistic literature.

A summer institute on nonparametric methods will be held at Minneapolis before this letter is read, and an advisory committee will report to the Council at this meeting on what plans, if any, should be made for a summer institute in 1959.

The final duty of the president is to appoint a new nominating committee, and I herewith appoint: T. E. Harris, Chairman, Herman Chernoff, David Cox, J. C. Kiefer, and W. J. Youden.

Deeply regretting having to take my leave thus, in absentia, I am

Most sincerely yours,
L. J. SAVAGE
President

IMS OFFICERS, COMMITTEES, AND REPRESENTATIVES FOR 1957–1958.

Council Members and Officers

Terms Expire 1958	*Terms Expire 1959*	*Terms Expire 1960*
R. C. Bose	T. W. Anderson	David Blackwell
Churchill Eisenhart	M. S. Bartlett	Harold Hotelling
Oscar Kempthorne	J. Berkson	Jerzy Neyman
W. J. Youden	E. L. Lehmann	I. R. Savage

President: L. J. Savage — Secretary: G. E. Nicholson
President-elect: J. Wolfowitz — Treasurer: A. H. Bowker
Editor: T. E. Harris

Fellows elected in 1958: Julius R. Blum, James Durbin, Benjamin Epstein, J. Hemelrijk, Leo Katz, Tatsuo Kawata, George E. Nicholson, Jr., Howard Raiffa, Sixto Rios, Stefan Vajda, Geoffrey S. Watson, Lionel Weiss.

Committees

(*The first person named is the chairman*)

IMS COMMITTEE ON EXCHANGES: P. S. Dwyer.

IMS COMMITTEE ON FELLOWS: Frank J. Anscombe, Z. W. Birnbaum, L. A. Goodman, W. Hoeffding, E. S. Pearson, E. L. Scott.

IMS FINANCE COMMITTEE: Mel Peisakoff, A. H. Bowker, Cuthbert Hurd, Theodore Yntema.

IMS MEMBERSHIP COMMITTEE: Benjamin Epstein, H. E. Daniels, Meyer Dwass, Solomon Kulback, Sigeiti Moriguti, G. R. Seth, Rosedith Sitgreaves, Milton Terry.

IMS COMMITTEE FOR INSTITUTIONAL MEMBERS: Mervin Muller, Z. W. Birnbaum, R. Bradford Murphy, Frank Akutowicz, K. J. Arnold, S. S. Wilks.

IMS COMMITTEE ON PROFESSIONAL STANDARDS: Joseph Lev.

IMS PROGRAM COMMITTEE FOR 1958 ANNUAL MEETING: W. Kruskal, F. J. Ans-

combe, R. J. Bose, D. L. Burkholder, D. G. Chapman, Cuthbert Daniel, T. S. Ferguson, Evelyn Fix, F. A. Graybill, H. O. Hartley, P. J. McCarthy, Howard Raiffa.

IMS PROGRAM COMMITTEE OF THE CENTRAL REGION: Jack Silber, Virgil Anderson, Charles Bell, H. T. David, F. A. Graybill, E. R. Immel, Bernard Ostle, M. B. Wilk (ex officio), W. Kruskal (ex officio), R. N. Bradt, Frank Graybill, Irving Burr, John Gurland, Howard Jones, Fred Leone, Boyard Rankin, Paul Rider, Jack Silbert (ex officio), Martin Wilk (ex officio).

IMS EASTERN REGION PROGRAM COMMITTEE: Boyd Harshbarger, Ralph Bradley, B. G. Greenberg, D. G. Horvitz, Carl Kossack, Herbert A. Meyer, John Pratt, Dorothy Gilford (ex officio), Martin Wilk (ex officio), W. H. Horton, J. D. Hromi.

IMS WESTERN REGION COMMITTEE: Richard Link, Fred Andrews, Charles Bell, Tom Ferguson, John Gilbert, John Hofmann, Marion Sandomire, David Stoller, Robert Tate, J. R. Vatnsdal, Frank Massey.

IMS COMMITTEE FOR SPECIAL INVITED PAPERS: W. S. Connor, G. E. P. Box, Kai-Lai Chung, W. Kruskal, G. E. Noether, M. B. Wilk, T. E. Harris member ex officio (as Editor).

IMS AD HOC COMMITTEE ON HIGH SPEED COMPUTING: A. S. Householder, G. S. Acton, R. L. Anderson, K. J. Arnold, C. F. Kossack, W. H. Kruskal, W. J. Merrill, H. A. Meyer, J. Moshman, H. W. Norton, G. J. Resnikoff, R. Slimak, Z. Szatrowski, D. Teichroew.

IMS SUBSCRIPTIONS COMMITTEE: Edward Coleman, Joe Adams, K. A. Bush, Lila Elveback, Harry Harman.

IMS COMMITTEE ON MATHEMATICAL TABLES: D. B. Owen, G. P. Steck, Paul Cox, Fred C. Leone, John W. Tukey, Marvin Zelen, R. L. Anderson, A. H. Bowker, E. E. Cureton, W. J. Dixon, C. W. Dunnett, Churchill Eisenhart, J. A. Greenwood, H. O. Hartley, William Kruskal, Daniel Teichroew, M. A. Woodbury, L. J. Savage (ex officio), J. Wolfowitz (ex officio).

IMS AD HOC COMMITTEE ON ANNUAL MEETING POLICY: Leo Katz, Cecil Craig, Churchill Eisenhart, Robert Hooke, Henry Scheffe, Martin Wilk.

AD HOC COMMITTEE TO INVESTIGATE THE POSSIBILITY OF BILLING FOR PUBLICATION IN THE *ANNALS*: A. M. Mood, A. H. Bowker, John Curtiss, Dorothy Gilford, W. Kruskal.

IMS BLACKBOARD COMMITTEE: Irving W. Burr, Herbert Robbins, Martin Wilk.

IMS EDITOR HUNT COMMITTEE: W. G. Cochran, T. W. Anderson, M. S. Bartlett, T. E. Harris, W. A. Wallis, S. S. Wilks.

IMS COMMITTEE ON RUSSIAN TRANSLATIONS: Ingram Olkin, Eugene Lukacs.

COMMITTEE TO ADVISE ON A POSSIBLE SUMMER INSTITUTE FOR 1959: David Blackwell, H. O. Hartley, Richard Savage, David Wallace, Max Woodbury.

IMS REPRESENTATIVE TO AAAS: Harold Hotelling.

AMERICAN STANDARDS ASSOCIATION COMMITTEE ON STATISTICAL NOMENCLATURE IMS REPRESENTATIVE FOR 1957–58: Max Halperin.

IMS REPRESENTATIVE IN DIVISION OF MATHEMATICS NATIONAL RESEARCH COUNCIL: W. Allen Wallis.

REPRESENTATIVES TO CONFERENCE ORGANIZATION OF THE MATHEMATICAL SCIENCES: Joseph F. Daly for 1957–58, H. B. Mann for 1957–59.

REPORT OF THE SECRETARY FOR 1958

During the past year The Institute has held its 75th through 78th meetings. A business meeting was held during the 78th (21st Annual) meeting. The Program Committees are to be congratulated on the excellent programs which

have been arranged under the immediate direction of David S. Stoller, Jack Silber, Herbert A. Meyer, and W. H. Kruskal with the overall guidance of our Program Coordinator, M. B. Wilk. The Assistant Secretaries, John F. Hofmann, Herbert T. David, Marvin Kastenbaum, and John W. Pratt, are to be congratulated on the physical arrangements, and the Associate Secretaries, Evelyn Fix, Jack Silber, and Dorothy M. Gilford, on their performance of the duties of the Secretary with respect to meetings.

MINUTES OF THE ANNUAL BUSINESS MEETING AUGUST 27, 1958

The annual business meeting of the Institute of Mathematical Statistics was called to order at 5:30 p.m. August 27, 1958, in Kresge Auditorium in Cambridge, Massachusetts by George E. Nicholson, Jr., Secretary, in the absence of the President and President-Elect. Approximately 45 members were present.

Minutes of the September 11, 1957 business meeting held in Atlantic City were approved.

Reports of the Secretary, Treasurer, Editor, Program Coordinator and President were presented and accepted.

Ballots were distributed to those members who had not voted by mail.

A resolution expressing appreciation of the work of the local arrangement committee and the staff of M.I.T. for arrangements for the meeting was read and passed.

G. E. Nicholson, Jr. announced that unless there were objections, W. H. Kruskal as an ex officio member of the Council would be ruled ineligible for election and that the four highest candidates after eliminating Kruskal would be ruled elected. It was moved, seconded and passed that the ruling be accepted.

The results of the election are as follows: J. W. Tukey, President-Elect; William Kruskal, Editor; T. E. Harris, Council Member, 1959–61; S. S. Wilks, Council Member, 1959–61; F. J. Anscombe, Council Member, 1959–61; Leo Katz, Council Member, 1959–61.

The meeting was adjourned about 6:30 P.M.

REPORT OF THE EDITOR FOR 1958

During the year ending July 31, the *Annals* received a larger number of new manuscripts than in any previous year (the total number of pages was slightly smaller than in the preceding year). The size of the printed volume for 1958 will be about 1300 pages, and this has been adequate to maintain the backlog of accepted, unprinted papers at substantially less than one issue. The Council has taken notice of the financial problems posed by a larger *Annals*, and plans are now being discussed for raising the required funds.

I take this opportunity to thank the many people whose hard work has been so valuable to the *Annals* during my editorial term, now ending. A list of persons who refereed papers during 1958 will be printed in an early issue. I thank particularly Ann Greene, Jeanette Hiebert, Dorothy Stewart, Helena Williams, and Margaret Wray, who have carried on the work of the editorial office, and I want to express my appreciation to The RAND Corporation for making possible my work on the *Annals*.

T. E. HARRIS
Editor

PUBLICATIONS RECEIVED

Introduccion a La Investigacion Operativa, Instituto de Investigaciones Estadisticas, Serrano 123, Madrid, Spain.

Anuario Estadistico de Espana, Presidencia del Gobierno, Instituto Nacional de Estadistica, Ferraz 41, Madrid, Spain.

Integrals of Airy Functions, National Bureau of Standards Applied Mathematics Series 52, issued May 15, 1958, 28 pp., 25 cents. (Order from the Superintendent of Documents, U.S. Government Printing Office, Washington 25, D. C.)

Table of Natural Logarithms for Arguments Between Five and Ten to Sixteen Decimal Places, National Bureau of Standards Applied Mathematics Series 53, issued March 29, 1958, $4.00. Supersedes Mathematical Table 12. (Order from Superintendent of Documents, U. S. Government Printing Office, Washington 25, D. C.)

Feinstein, Amiel, *Foundations of Information Theory*, McGraw-Hill Book Company, 330 West 42nd Street, New York 36, New York. 137 pages, $6.50.

Contents

INNEHALL

ARTICLES

METRON

Contents

Vol. XIX - N. 1-2 **31-VII-1958**

SOMMARIO—SOMMAIRE—CONTENTS—INHALT—SUMARIO

PAGE

Manuscripts submitted for publication and applications of subscribers as well as the sums for the subscription are to be addressed to Prof. CORRADO GINI "*Amministrazione del Metron*" *via Terme Diocleziane 10—ROME—(Italy).*

TRABAJOS DE ESTADISTICA

Review published by "Instituto de Investigaciones Estadísticas" of the "Consejo Superior de Investagaciones Científicas." Madrid, Spain.

Vol. IX **CONTENTS** **Cuaderno II**

For everything in connection with works, exchanges and subscription write to Professor Sixto Ríos, Instituto de Investigaciones Estadísticas, Consejo Superior de Investigaciones Científicas (Serrano, 123). Madrid, Spain. The Review is composed of three fascicles published three times a year (about 350 pages), and its annual price is 100 pesetas for Spain and South America and $4.00 U.S.A. for all other countries.